END

FILMED

DTIC

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR. 85-0688	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Stochastic Adaptive Control and Estimation Enhancement		5. TYPE OF REPORT & PERIOD COVERED Final Report Feb. 1980 – April 1984
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Y. Bar-Shalom		8. CONTRACT OR GRANT NUMBER(s) AFOSR-80-0098
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Connecticut, U-157 Storrs, CT 06268		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F, 2304, A1
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR Mathematical and Information Sciences Bolling AFB, DC 20332		12. REPORT DATE 3-19-85
		13. NUMBER OF PAGES 71
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Unrestricted

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stochastic Control	Multiple Model Control
Adaptive Control	Estimation Enhancement
Dual Effect	Parameter Identification Enhancement
Dual Controller	Cautious Control
	Probing Control

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This investigation dealt with the control of stochastic systems with algorithms that utilize the current information and its quality as well as the anticipated quality of future information. Such algorithms have the capability of probing the system in order to enhance the estimation or identification of its unknown parameters. Several algorithms were obtained for both continuous and discrete-valued uncertainties.

85 9 10 037

AD-A159 053

DTIC FILE COPY

1.   Introduction

This report presents the results of the investigations conducted over a period of four years on control algorithms designed for stochastic systems. The main feature of these algorithms is that they account for *these*:

(i) 1) the current uncertainty in the system, *and*

(ii) 2) the anticipated future uncertainty in the system, which is, in general control-dependent

The first feature leads to the control to have the "cautious" property in order to minimize the effect of the current uncertainties on the system's performance.

The second feature allows the control to affect in addition to the system's state also the system's uncertainty. Such a controller is called "dual controller" because, by taking advantage of its "dual effect" has the capability of reducing the future uncertainties.

These uncertainties can pertain to the system's state or its unknown parameters. Both continuous-valued and discrete-valued uncertainties have been considered.

The next section summarizes the major results of the research effort that have been published in reading control journals and presented at major national and international conferences.

2. Summary of Results

In the following an outline of each publication is given. The full papers appear in the Appendix.

2.1. C.J. Wenk and Y. Bar-Shalom, "A Multiple Model Adaptive Dual Control Algorithm for Stochastic Systems with Unknown Parameters," IEEE Trans. Automatic Control, vol. AC-25, pp. 703-710, Aug. 1980.

In this work an adaptive dual control algorithm is presented for linear stochastic systems with constant but unknown parameters. The system parameters are assumed to belong to a finite set on which a prior probability distribution is available. The tool used to derive the algorithm is preposterior analysis: a probabilistic characterization of the future adaptation process allows the contrtoller to take advantage of the dual effect. The resulting actively adaptive control called model adaptive dual (MAD) control is compared to two passively adaptive control algorithms-the heuristic certainty equivalence (HCE) and the De-shpande-Upadhyay-Lainiotis (DUL) model-weighted controllers. An analysis technique developed for the comparison of different controllers is used to show statistically significant improvement in the performance of the MAD algorithm over those of the HCE and DUL.

2.2 Y. Bar-Shalom, "Stochastic Dynamic Programming, Caution and Probing," IEEE Trans. Automatic Control, vol. AC-26, pp. 1184-1195, Oct. 1981.

The purpose of this paper is to unify the concepts of caution and probing put forth by Feldbaum with the mathematical technique of stochastic dynamic programming originated by Bellman. The recently developed decomposition of the expected cost in a stochastic control problem, is used to assess quantitatively the caution and probing effects of the system, uncertainties on the control. It is shown how in some problems, because of the

uncertainties, the control becomes cautious (less aggressive) while in

other problems it will probe (by becoming more aggressive) in order to

enhance the estimation/identification while controlling the system.

Following this a classification of stochastic control problems

according to the dominant effect is discussed. Tnis is then used to point

out which are the stochastic control problems where substantial improve-

ments can be expected from using a sophisticated algorithm versus a

simple one.

2.3 Y. Bar-Shalom and J. A. Molusis, "Stochastic Control and Identification Enhancement for the Flutter Suppression Problem," *Proc. 8th IFAC World Congress*, Kyoto, Japan, Aug. 1981.

The topic of this paper is the application of some recent results in

Stochastic control to an aerospace problem where there are large

uncertainties in the dynamics of the plant to be controlled. An approxi-

mation to the stochastic Dynamic Programming is considered that results

in an adaptive control of the "closed-loop" type: it utilizes feedback

(latest state and parameter estimates and their uncertainties) as well

as their anticipated future uncertainties - it anticipates (subject to

causality) subsequent feedback. This algorithm has the feature that

allows the control to enhance the parameter identification in real time.

This is done using the control's dual effect: the control can affect

the state as well as the (augmented) state uncertainty and thus can

reduce the uncertainty about some parameters. A flight control applica-

tion in which stochastic adaptive control appears to offer significant

payoff is the active control of aircraft wing-store flutter. Improved

flutter suppression can be accomplished with an adaptive controller that

has the capability to learn and identify the flutter modes during the

flight.

2.4   C. J. Wenk and Y. Bar-Shalom, "Model Adaptive Dual Control of MIMO Stochastic Systems," Proc. 20th IEEE Conf. on Decision and Control, San Diego, CA, Dec. 1981.

An adaptive dual control algorithm is presented for multiple-input,

multiple output (MIMO) linear systems with input and output noise and

unknown parameters. The system parameters are assumed to belong to a

finite set on which a prior probability distribution is available. The

difficulties in characterizing the future evolution of the MIMO system

information as required by the dynamic programmig are overcome through

a novel way of using preposterior analysis. This provides a proba-

bilistic characterization of the future adaptation process and allows

the controller to take advantage of the dual effect.

2.5   Y. Bar-Shalom, P. Mookerjee and J. A. Molusis, "A Linear Feedback Dual Controller for a Class of Stochastic Systems," Proc. CNRS Colloq. Automatique, Belle-Ile, France, Sept. 1982.

The methodology for deriving a dual control algorithm that has a linear

feedback form is presented. This control, while simple, has the capa-

bility of enhancing the identification of the system's unknown

parameters. A dual controller for a plant describing the helicopter

higher harmonic vibration control problem is presented together with

simulation results.

2.6   K. Birmiwal and Y. Bar-Shalom, "Dual Control Guidance for Simultaneous Identification and Interception of a Target," Automatica 20:737-749, Nove. 1984.

An adaptive dual-control guidance algorithm is presented for intercepting

a moving target in the presence of an interferring target (decoy) in a

stochastic environment. Two sequences of measurements are obtained at

discrete points in time; however, it is not certain which sequence came

from the target of interest and which from the decoy. Associated with

each track, the interceptor also receives noisy, state-dependent feature
measurements. The optimum control for the interceptor which is given by
the solution of the stochastic dynamic programming equation is not numeri-
cally feasible to obtain. An approximate solution of this equation is
obtained by evaluating the value of the future information gathering.
This is done through the use of preposterior analysis:approximate prior
probability densities are obtained and used to describe the future
learning and control. In this way, the interceptor control is used
for information gathering in order to reduce the future target and decoy
decoy inertial measurement errors and enhance the observable target/decoy
feature differences for subsequent discrimination between the true target
and the decoy. Simulation studies have shown the effectiveness of the
scheme.

2.7 J. A. Molusis, P. Mookerjee and Y. Bar-Shalom, "Dual Adaptive Control
Based upon Sensitivity Functions," Proc. 23rd IEEE Conf. on Decision
and Control, Las Vegas, NV, Dec. 1984.

A new adaptive dual control solution is presented for the control of a
class of multi-variable input-output system. Both rapidly varying
random parameters and constant but unknown parameters are included. The
new controller modifies the cautious control design by numerator and
denominator correction terms. This controller is shown to depend upon
sensitivity functions of the expected future cost. A scalar example
is presented to provide insight into the properties of the new dual
controller. Monte-Carlo simulations are performed which show improve-
ment over the cautious controller and the Linear Feedback Dual Con-
troller.

Appendix

A Multiple Model Adaptive Dual Control Algorithm
for Stochastic Systems with Unknown Parameters

CARL J. WENK, STUDENT MEMBER, IEEE, AND YAAKOV
BAR-SHALOM, SENIOR MEMBER, IEEE

Abstract—An adaptive dual control algorithm is presented for linear stochastic systems with constant but unknown parameters. The system parameters are assumed to belong to a finite set on which a prior probability distribution is available. The tool used to derive the algorithm is preposterior analysis: a probabilistic characterization of the future adaptation process allows the controller to take advantage of the dual effect. The resulting actively adaptive control called model adaptive dual (MAD) control is compared to two passively adaptive control algorithms—the heuristic certainty equivalence (HCE) and the De-shpande-Upadhyay-Lainiotis (DUL) model-weighted controllers. An analysis technique developed for the comparison of different controllers is used to show statistically significant improvement in the performance of the MAD algorithm over those of the HCE and DUL.

I. INTRODUCTION

In the control of linear stochastic systems with quadratic cost, the certainty equivalence property [6] is known to hold. If, however, there are unknown parameters in the system to be controlled, then certainty equivalence does not hold and the dynamic programming cannot be solved [1]. In this case a control decision is known to affect not just the future state of the system, but also the future state and parameter uncertainty; that is, the control has the dual effect, first discussed by Feldbaum [15], and later shown to be intimately related to the certainty equivalence property [6].

Because the parameter uncertainty renders the optimum control solution unattainable, a number of parameter-adaptive suboptimum control strategies have been sought [14],[19] [23],[13],[2],[24]. With the exception of [24], most of these strategies, however, are in the passive feedback classification as discussed in [6]. That is, they do not take into account the knowledge that future learning about the unknown system parameters will occur. An algorithm which uses such knowledge to improve its control decisions is called actively adaptive; it takes advantage of the dual effect of the control to improve the identification and ultimately the performance.

This paper presents an actively adaptive control algorithm for linear stochastic systems where the vector θ consisting of the constant but unknown system parameters, is equal to one of the M known model parameter vectors $\theta_j, j = 1, \cdots, M$. This assumption that the true system is a member of a discrete set of known model systems has been used for the development of a number of passively adaptive control algorithms [13],[19],[21],[23]. and has received considerable recent attention in, for example, the adaptive flight control problem for the F-8 Digital-Fly-By-Wire Aircraft [2],[3] Performance difficulties have arisen, however, due to the inherently passive learning properties of existing algorithms designed for the multiple model adaptive control problem The algorithm presented in this paper, with its active learning properties, should represent an advance toward a more sophisticated solution of the multiple model problem.

The actively adaptive control algorithm presented here, called the model adaptive dual (MAD) control algorithm, is developed and studied within the context of controlling the output of a single-input, single-output system, in order to help gain understanding of the dual effect of the control in the multiple model problem. The problem is formulated in Section II. The MAD algorithm for two models is obtained in Sections III–V by constructing an approximate solution to the stochastic dynamic programming, the exact solution of which would give the globally optimum (dual) control. Evaluation of the value of future information gathering will be made through the use of preposterior analysis [18]; approximate prior probability densities are obtained and used to describe future learning and control. The extension to $M > 2$ models is presented in Section VI.

Numerical studies and comparisons of the MAD algorithm are made in Section VII with two passive algorithms, the heuristic certainty equivalence (HCE) algorithm, and the Deshpande-Upadhyay-Lainiotis (DUL) algorithm, as well as with the optimal controls produced for each model system with *known* parameters. A rigorous statistical analysis technique is presented for a meaningful comparison of the performances obtained from Monte Carlo simulations employing the above algorithms.[1] It is shown by statistical tests performed on the results of a Monte Carlo simulation procedure that significant performance improvements may be achieved using MAD over HCE and DUL. In the latter algorithm, used in the F-8 aircraft problem in [2], the control is formed as a weighted sum of the model-optimal controls.

Lastly, while the MAD algorithm is designed for eventual on-line computational feasibility, it is more expensive than HCE and DUL. It is also pointed out that MAD has a built-in feature to help determine *a priori*, in a non-Monte-Carlo fashion, when the performance improvements obtainable with MAD are large enough to warrant the added computing load.

II. Problem Formulation

Consider controlling the linear system described by the input–output model [4]

$$v(t) = A(q^{-1})v(t-1) + B(q^{-1})u(t-1) + e(t) \quad (2.1)$$

where

$$A(q^{-1}) = a_1 + a_2 q^{-1} + \cdots + a_n q^{-(n-1)} \quad (2.2)$$

$$B(q^{-1}) = b_1 + b_2 q^{-1} + \cdots + b_n q^{-(n-1)} \quad (2.3)$$

---

[1] To the best knowledge of the authors, past comparisons between different control algorithms were limited to sample means, leaving open the question of statistical significance of the observed differences.

are polynomials in the delay operator q^{-1} defined by $q^{-1}z(t) = z(t-1)$. The system output is $y(t)$, the input is $u(t)$, and $e(t)$ is a zero-mean, white Gaussian disturbance with standard deviation λ. Part or all of the parameter vector defined by

$$\theta^T = [a_1 a_2 \cdots a_n b_1 b_2 \cdots b_n] \quad (2.4)$$

is unknown. It is assumed, however, that the true parameter vector θ is equal to one of M known constant model vectors $\theta_j, j = 1, \cdots, M$, with corresponding known *a priori* probabilities

$$P[\theta = \theta_j] = \Lambda_j(0); \quad j = 1, \cdots, M \quad (2.5)$$

$$\sum_{j=1}^{M} \Lambda_j(0) = 1 \quad (2.6)$$

The objective is to determine a sequence of control decisions $\{u(0), u(1), \cdots, u(N-1)\}$ which minimizes

$$J(0) = E[C(0)] \quad (2.7)$$

where the cost is quadratic about a reference trajectory

$$C(t) = \frac{1}{2} q(N)[y(N) - y_r(N)]^2$$

$$+ \frac{1}{2} \sum_{\tau=t}^{N-1} \{ q(\tau)[y(\tau) - y_r(\tau)]^2 + r(\tau)[u(\tau) - u_r(\tau)]^2 \} \quad (2.8)$$

subject to (2.1)–(2.6). The expectation in (2.7) is performed with respect to all random variables in $C(0)$, with the quantities $q(t), r(t), y_r(t)$, and $u_r(t)$ all known (time-varying) constants, $t = 0, 1, \cdots, N$. The information vector at time t, $Z(t)$ consists of the sequence of known outputs and control decisions

$$Z(t) = \{ y(0), y(1), \cdots, y(t), u(0), u(1), \cdots, u(t-1) \}. \quad (2.9)$$

Given that an admissible control decision $u(t)$ is a function of $Z(t)$ as well as the statistical description of the future observations [6], the optimum solution to the problem is given by the stochastic dynamic programming as

$$u^*(t) = \arg \min E \left\{ \frac{1}{2} q(t)[y(t) - y_r(t)]^2 + \frac{1}{2} r(t)[u(t) - u_r(t)]^2 \right.$$

$$\left. + J^*[t+1, u(t)] | Z(t), u(t) \right\} \quad (2.10)$$

where $J^*[t+1, u(t)]$ is the optimum cost-to-go from $t+1$ to the end, and is a function of the present control decision $u(t)$. The globally optimum control cannot, in general, be computed—the only sure way of avoiding the "curse-of-dimensionality" [11] is by finding a recursion in the cost-to-go, which here does not exist because of the parameter uncertainty. Several computable suboptimal control algorithms for this problem do exist, however, including two of particular interest here. They are the so-called heuristic certainty equivalence (HCE) algorithm [6], and the Deshpande-Upadhyay-Lainiotis (DUL) algorithm [13]. In the HCE algorithm, a current best estimate of θ is computed as

$$\hat{\theta}(t) = \sum_{j=1}^{M} \Lambda_j(t)\theta_j. \quad (2.11)$$

$\hat{\theta}(t)$ is then used as if it were the true parameter vector, under which assumption the optimum control is easily computed. Thus, in a heuristic manner, certainty equivalence (though untrue) is enforced. In the DUL algorithm, the control decision is obtained as

$$u(t) = \sum_{j=1}^{M} \Lambda_j(t)u_j(t) \quad (2.12)$$

where $u_j(t)$ is the optimum control which would result if $\theta = \theta_j$ were in fact true (again easily computed). Both the HCE and DUL algorithms are passively adaptive; they do not assess the effect which the current control decision will have on future learning. HCE and DUL are algorithms of the feedback type, rather than of the truly closed-loop type as defined in [6]. The optimum control to be derived by approximation

of (2.10) is a closed-loop control, capable of taking advantage of the dual effect [7] of the control in this problem.

For the moment, attention is focused on the two-model $(M=2)$ case since, as will be shown in Section VI, solution of the general M-model case $(M \geqslant 2)$ may be obtained by solving a fixed set of two-model subproblems. The problem of pairwise $(M=2)$ model discriminations will be shown to embody the basic duality of the control.

III. PREPOSTERIOR ANALYSIS AND THE APPROXIMATE SOLUTION OF THE STOCHASTIC DYNAMIC PROGRAMMING EQUATION

Consider the case $M=2$ with the probabilities defined in (2.5)

$$\Lambda_1(0) = \Pi(0), \qquad \Lambda_2(0) = 1 - \Pi(0) \tag{3.1}$$

where the prior $\Pi(0)$ is known.

In order to obtain a computationally implementable algorithm, the cost-to-go in (2.10) will be approximated as follows: the future controls (for $i > t+1$) will be assumed of fixed structure; they will be of the DUL type but with time-varying probabilities as more information becomes available to the controller. This is expressed as follows:

$$E\{J^*(t+1)|Z(t),u(t)\}$$
$$\cong E\left\{ \min_{L(t+1)} E[C(t+1)|Z(t+1),L(t+1)]|Z(t),u(t)\right\} \tag{3.2}$$

where $L(t+1)$ is the set of all parameters in the controller structure from $t+1$ through the end. Using the total probability theorem, the optimum cost-to-go in (3.2) may be written as

$$J^*(t+1) \approx \min_{L(t+1)} \{\Pi(t+1)E[C(t+1)|Z(t+1),L(t+1),\theta=\theta_1]$$
$$+ [1-\Pi(t+1)]E[C(t+1)|Z(t+1),L(t+1),\theta=\theta_2]\} \tag{3.3}$$

where

$$\Pi(t+1) = P[\theta=\theta_1|Z(t+1)] \tag{3.4}$$

is given by Bayes' rule

$$\Pi(t+1) = \frac{p[y(t+1)|Z(t),u(t),\theta=\theta_1]\Pi(t)}{p[y(t+1)|Z(t),u(t)]}$$

$$= \frac{p[y(t+1)|Z(t),u(t),\theta=\theta_1]\Pi(t)}{\Pi(t)p[y(t+1)|Z(t),u(t),\theta=\theta_1]+[1-\Pi(t)]p[y(t+1)|Z(t),u(t),\theta=\theta_2]} \tag{3.5}$$

where

$$p[y(t+1)|Z(t),u(t),\theta=\theta_j] = \frac{1}{\sqrt{2\pi}\lambda}\exp\left\{-\frac{1}{2\lambda^2}[y(t+1)-\hat{y}_j(t+1)]^2\right\} \tag{3.6}$$

$$\hat{y}_j(t+1) = E[y(t+1)|Z(t),u(t),\theta=\theta_j] \tag{3.7}$$

for $j=1,2$. If A_j, B_j denote the polynomials of (2.2) and (2.3), respectively, formed assuming $\theta=\theta_j$ is true, then (3.7) becomes

$$\hat{y}_j(t+1) = A_j y(t) + B_j u(t) \tag{3.8}$$

From (3.5) one can obtain the inverse transformation from $\Pi(t+1)$ to the latest observation $y(t+1)$

$$y(t+1) = \frac{1}{2}[\hat{y}_1(t+1)+\hat{y}_2(t+1)]$$
$$+ \frac{\lambda^2}{\hat{y}_2(t+1)-\hat{y}_1(t+1)} \ln\left\{\frac{\Pi(t)[1-\Pi(t+1)]}{[1-\Pi(t)]\Pi(t+1)}\right\} \tag{3.9}$$

Thus, the outer expectation on the right-hand side of (3.2), which is over $y(t+1)$, can be replaced by an expectation over $\Pi(t+1)$ as follows:

$$E[J^*(t+1)|Z(t),u(t)]$$
$$= \int_0^1 \min_{L(t+1)} \{\Pi(t+1)E[C(t+1)|Z(t+1),L(t+1),\theta=\theta_1]$$
$$+ [1-\Pi(t+1)]E[C(t+1)|Z(t+1),L(t+1),\theta=\theta_2]\}$$
$$\cdot p[\Pi(t+1)|Z(t),u(t)]d\Pi(t+1) \tag{3.10}$$

where $p[\Pi(t+1)|Z(t),u(t)]$ is the preposterior probability density function of $\Pi(t+1)$, which is the information state for the parameters at $t+1$. The term preposterior [18] means that this is the *prior* density (with respect to time $t+1$) of the *posterior* $\Pi(t+1)$, conditioned on the information at t. This density is obtained using (3.9) as

$$p[\Pi(t+1)|Z(t),u(t)] = \left|\frac{\partial y[\Pi(t+1)]}{\partial\Pi(t+1)}\right| p[y(t+1)|Z(t),u(t)]$$

$$= \frac{\lambda^2}{|\hat{y}_1(t+1)-\hat{y}_2(t+1)|}\frac{1}{\Pi(t+1)[1-\Pi(t+1)]}\left\{\Pi(t)\exp\right.$$

$$\left[-\frac{1}{2\lambda^2}\left[\frac{\lambda^2}{\hat{y}_2(t+1)-\hat{y}_1(t+1)}\ln\frac{\Pi(t)[1-\Pi(t+1)]}{[1-\Pi(t)]\Pi(t+1)}\right.\right.$$

$$\left.\left.-\frac{1}{2}(\hat{y}_1(t+1)-\hat{y}_2(t+1))\right]^2\right]+[1-\Pi(t)]\exp$$

$$\left[-\frac{1}{2\lambda^2}\left[\frac{\lambda^2}{\hat{y}_2(t+1)-\hat{y}_1(t+1)}\ln\frac{\Pi(t)[1-\Pi(t+1)]}{[1-\Pi(t)]\Pi(t+1)}\right.\right.$$

$$\left.\left.+\frac{1}{2}(\hat{y}_1(t+1)-\hat{y}_2(t+1))\right]^2\right]\right\} \tag{3.11}$$

The integration required in (3.10) is still not feasible to perform, even given knowledge of the exact preposterior density (3.11). An approximate solution to this integration is obtained by taking advantage of a fundamental property of the preposterior density: as the signal-to-noise ratio

$$SNR = \frac{1}{\lambda^2}[\hat{y}_1(t+1)-\hat{y}_2(t+1)]^2 \tag{3.12}$$

increases, the ability to discriminate between the two models increases, and the preposterior density in (3.11) exhibits a distinct bimodal character (see [9]); most of the density becomes concentrated around two distinct locations, say $\Pi_1(t+1)$ and $\Pi_2(t+1)$. In the limit as $SNR \to \infty$, $p[\Pi(t+1)|Z(t),u(t)]$ becomes the weighted sum of two delta functions

$$\lim_{SNR\to\infty} p[\Pi(t+1)|Z(t),u(t)]$$
$$= \Pi(t)\delta[\Pi(t+1)-1]+[1-\Pi(t)]\delta[\Pi(t+1)]. \tag{3.13}$$

These observations suggest using the following approximate preposterior density.

$$p[\Pi(t+1)|Z(t),u(t)] \approx \Pi(t)\delta[\Pi(t+1)-\Pi_1(t+1)]$$
$$+ [1-\Pi(t)]\delta[\Pi(t+1)-\Pi_2(t+1)] \tag{3.14}$$

where the delta function locations $\Pi_1(t+1)$ and $\Pi_2(t+1)$ satisfy

$$0 \leqslant \Pi_2(t+1) \leqslant \Pi(t) \leqslant \Pi_1(t+1) \leqslant 1. \tag{3.15}$$

The locations $\Pi_1(t+1)$ and $\Pi_2(t+1)$ may be obtained by moment matching: they are chosen so that the first two moments of $\Pi(t+1)$ produced by the approximate density (3.14) match those of the true density (3.11). Such a technique has been used with success in [8],[9]. A simple and accurate technique to carry this out is described in Appendix I.

While the approximate preposterior density has now been established, evaluation of the cost-to-go in (3.10) still requires a minimization with respect to the set $L(t+1)$ of (time-varying) controller parameters from $i=t+1$ to the end of the control period, a set which, of course, depends on the statistic $\Pi(t+1)$. An approximate solution to the minimization in

(3.10), which is easy to implement, is obtained by assuming a future sequence of DUL type controls represented by $\bar{L}(t+1)$:

$$E[J^*(t+1)|Z(t),u(t)] \approx \hat{J}[t+1,u(t)]$$

$$= \int_0^1 \Big\{ \Pi(t+1)E\Big[C(t+1)|Z(t+1),\bar{L}(t+1),\theta=\theta_1 \Big]$$

$$+[1-\Pi(t+1)]E\Big[C(t+1)|Z(t+1),\bar{L}(t+1),\theta=\theta_2 \Big] \Big\}$$

$$\cdot \{\Pi(t)\delta[\Pi(t+1)-\Pi_1(t+1)]+[1-\Pi(t)]$$

$$\cdot \delta[\Pi(t+1)-\Pi_2(t+1)]\} \, d\Pi(t+1). \tag{3.16}$$

Performing the integration, (3.16) becomes

$$\hat{J}[t+1,u(t)] = \Pi(t)\Pi_1(t+1)\hat{J}_{11}\Big[t+1,u(t),\bar{L}_{11}(t+1),\theta=\theta_1 \Big]$$

$$+\Pi(t)[1-\Pi_1(t+1)]\hat{J}_{12}\Big[t+1,u(t),L_{12}(t+1),\theta=\theta_2 \Big]$$

$$+[1-\Pi(t)]\Pi_2(t+1)\hat{J}_{21}\Big[t+1,u(t),\bar{L}_{21}(t+1),\theta=\theta_1 \Big]$$

$$+[1-\Pi(t)][1-\Pi_2(t+1)]\hat{J}_{22}\Big[t+1,u(t),\bar{L}_{22}(t+1),\theta=\theta_2 \Big]. \tag{3.17}$$

Equation (3.17) represents the approximate cost-to-go resulting from a particular control choice $u(t)$. The nominal sequence of control parameters $\bar{L}_{ij}(t+1)$, $i,j=1,2$ consists of a DUL weighted sum of model control gains. This sum is computed with nominal weighting factors given by:

1) $\Pi(t+1)=\Pi_i(t+1)$ as the initial sufficient statistic for θ at $t+1$,

2) subsequent nominal posterior probabilities $\Omega_{ij}(i)$ that $\theta=\theta_1$ which evolve as $i=t+2,\cdots,N-1$ when this DUL control is applied to the system with $\theta=\theta_j$.

Note that the model control gains are obtained from a standard linear quadratic problem with known parameters. The term $\hat{J}_{ij}$, which is the corresponding cost, is obtained from a standard recursion for a known linear system with $\theta=\theta_j$, quadratic cost, and a given set of control parameters $\bar{L}_{ij}(t+1)$. See, for example, [10].

IV. THE NOMINAL SEQUENCE OF FUTURE POSTERIOR PROBABILITIES

The nominal future posterior probabilities $\Omega_{ij}(i)$ are generated by constructing a future observation and control scenario, based on the statistical information contained in the approximate preposterior density function (3.14). This density indicates that, given a specific control decision $u(t)$, with probability $\Pi(t)$ the posterior probability $\Pi(t+1)$ will become $\Pi_1(t+1)$, and with probability $[1-\Pi(t)]$ the posterior will become $\Pi_2(t+1)$. Using (3.9) it follows that the observation which would produce the posterior $\Pi(t+1)=\Pi_i(t+1)$, $i=1,2$, is given by

$$\bar{y}_i(t+1) = \frac{1}{2}[\dot{y}_1(t+1)+\dot{y}_2(t+1)] + \frac{\lambda^2}{\dot{y}_2(t+1)-\dot{y}_1(t+1)}$$

$$\cdot \ln\left\{ \frac{\Pi(t)[1-\Pi_i(t+1)]}{[1-\Pi(t)]\Pi_i(t+1)} \right\}. \tag{4.1}$$

The terms $\Pi_i(t+1)$, $\bar{y}_i(t+1)$ are now used as initial conditions at $i=t+1$ for a nominal future observation and control sequence; nominal outputs $\bar{y}_{ij}(i)$ for a given pair (i,j) are generated by replacing $e(i)$ by its mean, which is zero, in (2.1) with $\theta=\theta_j$:

$$\bar{y}_{ij}(i+1) = A_j\bar{y}_{ij}(i) + B_j\bar{u}_{ij}(i); \quad i=t+1,\cdots,N-2; \ \bar{y}_{ij}(t+1)=\bar{y}_i(t+1),$$

$$j=1,2 \tag{4.2}$$

The nominal controls $\bar{u}_{ij}(t)$ are generated using a DUL control policy

$$\bar{u}_{ij}(i) = \bar{a}_{ij}^T(i)\bar{x}_{ij}(i) + \bar{\beta}_{ij}(i); \quad i=t+1,\cdots,N-1 \tag{4.3}$$

where x is a suitable state vector corresponding to (2.1) and $\bar{x}_{ij}$ represents the future nominal state corresponding to $\bar{y}_{ij}$, which was specified above. The set of control parameters is

$$\bar{L}_{ij}(t+1) = \Big\{ \bar{a}_{ij}(i),\bar{\beta}_{ij}(i), i=t+1,\cdots,N-1 \Big\}. \tag{4.4}$$

The control gains are given by the weighted sums [13]

$$\bar{a}_{ij}(i) = \Omega_{ij}(i)\alpha_1(i) + [1-\Omega_{ij}(i)]\alpha_2(i) \tag{4.5}$$

where α_1,α_2 are the model-optimal gains. An analogous equation yields $\bar{\beta}_{ij}$.

The nominal posterior probability that the controller will attach to the parameter being $\theta=\theta_1$, when in fact it is $\theta=\theta_j$, $j=1$ or 2, and at $t+1$ it started with $\Pi_i(t+1)$, $l=1$ or 2, is, using (3.5)

$$\Omega_{ij}(i+1) = \left\{ 1 + \frac{1-\Omega_{ij}(i)}{\Omega_{ij}(i)} \exp\left\{ \frac{(-1)^j}{2\lambda^2}\Big[\bar{y}_{ij}(i+1)-\bar{\eta}_{ij}(i+1) \Big]^2 \right\} \right\}^{-1}$$

$$\tag{4.6}$$

$$i=t+1,\cdots,N-2; \quad \Omega_{ij}(t+1)=\Pi_i(t+1)$$

where the "mismatched" $(k\neq j)$ prediction is

$$\bar{\eta}_{ij}(i+1) \doteq E\Big[\bar{y}_{ij}(i+1)|\bar{Z}_{ij}(i),\bar{u}_{ij}(i),\theta=\theta_k \Big]$$

$$= A_k\bar{y}_{ij}(i) + B_k\bar{u}_{ij}(i); \quad k\neq j \tag{4.7}$$

with $\bar{Z}_{ij}(i)$ the nominal information vector at time i.

Equation (4.6) specifies the four "learning curves" used to compute the cost-to-go (3.17) in order to obtain a feasible solution to the stochastic dynamic programming equation (2.10). Note that if $j=1$ then Ω_{ij} converges toward unity for both $l=1$ and 2; however, because of (3.15) it will converge faster if $l=1$. Conversely, if $j=2$, then Ω_{ij} converges toward zero, again for both $l=1,2$, but faster if $l=2$.

V. SOME REMARKS ON THE PROPERTIES OF THE NEW ALGORITHM

From (2.9), (2.10), (3.16), and (3.17) it can be seen that the MAD control at time t is obtained by numerically locating a minimum with respect to $u(t)$ of the cost function

$$\hat{J}[t,u(t)] = \frac{1}{2}r(t)[u(t)-u_r(t)]^2 + \hat{J}[t+1,u(t)] \tag{5.1}$$

where $\hat{J}[t+1,u(t)]$ is the approximate cost-to-go as given by (3.17). A golden section line search combined with a quadratic fit [5] may be used to locate $u^{MAD}(t)$, where the HCE and DUL algorithm controls are used to set the initial control search window. Computational evidence indicates that between 5 and 8 function evaluations $\hat{J}[t+1,u(t)]$ (5–8 different values of $u(t)$) are sufficient to achieve high accuracy in locating the minimum.

By using the approximate preposterior density (3.14), consideration of the possible values $\Pi(t+1)$ may take on is reduced to two "most crucial" values, $\Pi_1(t+1)$ and $\Pi_2(t+1)$. Equation (3.17) indicates then that four possible events need to be considered: $\Pi_1(t+1)$ becomes the posterior with the true system $\theta=\theta_1$; $\Pi_1(t+1)$ becomes the posterior but $\theta=\theta_2$ is true; $\Pi_2(t+1)$ is the statistic but $\theta=\theta_1$; and $\Pi_2(t+1)$ occurs with the true system $\theta=\theta_2$. The probabilities of these four events occurring are $\Pi(t)\Pi_1(t+1)$, $\Pi(t)[1-\Pi_1(t+1)]$, $[1-\Pi(t)]\Pi_2(t+1)$, and $[1-\Pi(t)][1-\Pi_2(t+1)]$, respectively. The cost which will be incurred if the event described by the $\{\Pi_i(t+1),\theta_j\}$ pair happens is $\hat{J}_{ij}[t+1,u(t),\bar{L}_{ij}(t+1),\theta=\theta_j]$.

Consider now how $\hat{J}_{ij}$ realistically represents the cost of such an event. First assume $l=j$; for example, take $l=j=1$. Due to the condition described by (3.15), the output $\bar{y}_1(t+1)$ given by (4.1) with $l=1$ which would produce this $\Pi_1(t+1)$ would more likely come from a system where $\theta=\theta_1$ were true. Since $j=1$ in $\hat{J}_{11}$ this represents convergence of $\Pi(t+1)=\Pi_1(t+1)>\Pi(t)$ in the right direction, which is toward unity. In the future nominal control scenario described in Section IV the probabilities $\Omega_{11}(t)$ will then converge steadily toward unity, since the mismatched predicted observation (4.7) appears as a negative exponent in (4.6). If $l=j=2$ the exponent in (4.6) is positive and $\Omega_{22}(i)$ converges to zero. Now consider what happens if $l\neq j$; for example, if $l=2,j=1$. The true system has $\theta=\theta_1$, but a nominal observation $\bar{y}_2(t+1)$ occurs which makes $\Pi(t+1)=\Pi_2(t+1)<\Pi(t)$; i.e., $\Pi(t+1)$ goes in the wrong direction. In the subsequent nominal control scenario the observations

707

$\bar{y}_{21}(i)$ come from the true system with $\theta = \theta_1$, and thus, of course, the posteriors $\Omega_{21}(i)$ will recover from the "bad" initial $\Omega_{21}(t+1) = \Pi_2(t+1)$, but only after some time. Meanwhile, the control gains $\bar{L}_{21}(t+1)$ have been closer to the optimum gains of the wrong system ($\theta = \theta_2$), thus accumulating an added cost represented by J_{21}. Similar statements may be made about the event $l = 1$, $j = 2$. Thus, the cross terms J_{lj}, $l \neq j$, represent the costs incurred if learning is degraded by bad observations at $t+1$. Correspondingly, it is expected that

$$J_{21} > J_{11} \quad \text{and} \quad J_{12} > J_{22} \tag{5.2}$$

on a range of control values containing the HCE, DUL, and MAD controls. Computational evidence in Section VII indicates that this is indeed so.

Thus, the algorithm MAD is sensitive to the anticipated rate of future learning, and, if needed, its present decision will affect that learning rate appropriately.

The evolution of the information about the system during the process, described in detail in the previous two sections, can also be summarized in pictorial form as in Fig. 1. The current probability $\Pi(t)$ evolves into one of the two values Π_1 or Π_2 from which four "learning" curves follow. These curves are labeled lj and they correspond to the four cost components from (3.17). This is the essence of the novel approach that yields the closed-loop [7] approximation of the stochastic dynamic programming presented here.

VI. THE GENERAL M-MODEL MAD CONTROL ALGORITHM

Extension of the two-model MAD algorithm described in Sections III–V to include the general case of M models, $M > 2$, is now discussed. It will be shown that the M-model MAD algorithm consists of performing two-model cost computations for each of the distinct pairs of models using the two-model MAD algorithm, along with one-model optimum cost computations for an appropriate adjustment.

To begin the development, consider first the case $M = 3$. Define W_1, W_2, and W_3 as the three mutually exclusive and exhaustive events $\theta = \theta_1$, $\theta = \theta_2$, and $\theta = \theta_3$ true, respectively. Then the mixed probability expression [25] $p[J, W_1 \cup W_2 \cup W_3]$, where J is a random variable, can be written as

$$p[J, W_1 \cup W_2 \cup W_3] = p[J, W_1 \cup W_2] + p[J, W_1 \cup W_3] + p[J, W_2 \cup W_3]$$
$$- p[J, W_1] - p[J, W_2] - p[J, W_3] = p[J] \tag{6.1}$$

where the union $W_k \cup W_j$ signifies the event that one of θ_k, θ_j is the true parameter vector, and where $W_1 \cup W_2 \cup W_3$ is the sure event (note that $W_1 \cap W_2 \cap W_3 = \varnothing$, the null set). Using (6.1) for a cost-to-go $J(t+1)$, one can write

$$p[J(t+1)|Z(t), u(t)] = [\Lambda_1(t) + \Lambda_2(t)] p[J(t+1)|Z(t), u(t), W_1 \cup W_2]$$
$$+ [\Lambda_1(t) + \Lambda_3(t)] p[J(t+1)|Z(t), u(t), W_1 \cup W_3]$$
$$+ [\Lambda_2(t) + \Lambda_3(t)] p[J(t+1)|Z(t), u(t), W_2 \cup W_3]$$
$$- \Lambda_1(t) p[J(t+1)|Z(t), u(t), W_1] - \Lambda_2(t) p[J(t+1)|Z(t), u(t), W_2]$$
$$- \Lambda_3(t) p[J(t+1)|Z(t), u(t), W_3]. \tag{6.2}$$

From (6.2) it follows that

$$E[J(t+1)|Z(t), u(t)] = [\Lambda_1(t) + \Lambda_2(t)] E[J(t+1)|Z(t), u(t), W_1 \cup W_2]$$
$$+ [\Lambda_1(t) + \Lambda_3(t)] E[J(t+1)|Z(t), u(t), W_1 \cup W_3]$$
$$+ [\Lambda_2(t) + \Lambda_3(t)] E[J(t+1)|Z(t), u(t), W_2 \cup W_3]$$
$$- \Lambda_1(t) E[J(t+1)|Z(t), u(t), W_1] - \Lambda_2(t) E[J(t+1)|Z(t), u(t), W_2]$$
$$- \Lambda_3(t) E[J(t+1)|Z(t), u(t), W_3]. \tag{6.3}$$

Now, for arbitrary $M > 2$ it can be shown (see, e.g., [17]) that

$$E[J(t+1)|Z(t), u(t)]$$
$$= \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} [\Lambda_i(t) + \Lambda_j(t)] E[J(t+1)|Z(t), u(t), W_i \cup W_j]$$
$$- (M-2) \sum_{j=1}^{M} \Lambda_j(t) E[J(t+1)|Z(t), u(t), W_j]. \tag{6.4}$$
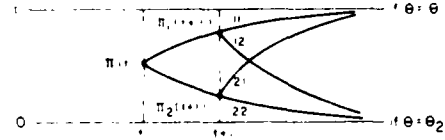
Equation (6.4) states the following.

Theorem: For a specified $u(t)$ the cost-to-go $J(t+1)$, given that one of M models θ_j, $j = 1, \cdots, M$ is correct, can be obtained as follows.

1) First compute the cost-to-go which results if one of either θ_k, θ_j, $k \neq j$, is true; this is done for each of $\bar{M} = M(M-1)/2$ distinct model pairs.

2) Compute the optimum model costs (θ_j true), $j = 1, \cdots, M$ and form the overall cost-to-go according to (6.4).

Of course, all the expectations in (6.4) are conditioned on the same information state $Z(t)$ and control choice $u(t)$. The model costs are easily computed from a standard linear quadratic problem. For each of the two-model costs $E[J(t+1)|Z(t), u(t), W_k \cup W_j]$ an approximate cost is computed using the MAD algorithm of Sections III–V. Since the event $W_k \cup W_j$ means that either θ_k or θ_j is true, the required sufficient statistic $\Pi^{kj}(t)$ in the two-model MAD cost evaluation for the specified pair of models is

$$\Pi^{kj}(t) = \frac{\Lambda_k(t)}{\Lambda_k(t) + \Lambda_j(t)} \tag{6.5}$$

thus maintaining proper normalization.

The general M-model ($M > 2$) MAD algorithm thus searches for a minimum in

$$\tilde{J}[t, u(t)] = \frac{1}{2} r(t) [u(t) - u_r(t)]^2 + \hat{J}[t+1, u(t)] \tag{6.6}$$

where $\hat{J}[t+1, u(t)]$ is given by (6.4).

VII. NUMERICAL EXAMPLES

In the numerical studies, attention was focused on studying the performance and characteristics of the MAD algorithm for the case $M = 2$, since the pairwise model discrimination procedure constitutes the very essence of the actively adaptive decision making process of the algorithm. Performance will be compared with that of the passively adaptive HCE and DUL algorithms.

Example 1: A second-order system ($n = 2$) is considered with two poles at 0.7. It is not certain whether the true system's zero is at -0.225 or at -0.9. Correspondingly, the true system parameter vector is one of the following:

$$\theta_1^T = [1.4 \quad -0.49 \quad 2 \quad 0.45] \tag{7.1}$$

$$\theta_2^T = [1.4 \quad -0.49 \quad 1 \quad 0.9] \tag{7.2}$$

which are considered *a priori* equiprobable. The initial output is $y(0) = 0.1$ and it is desired to make it follow over $N = 5$ time steps the reference trajectory (for $t = 0, 1, \cdots, 5$)

$$y_r = [0.1 \quad 0.5 \quad 1 \quad 2 \quad 2.5 \quad 10]. \tag{7.3}$$

The corresponding weightings in the cost (2.8) for $i = 0, 1, \cdots, 5$ were chosen as

$$q = [0 \quad 1 \quad 2 \quad 3 \quad 5 \quad 50]. \tag{7.4}$$

No penalty was attached to the control. Note that this would be a straightforward minimum variance controller about a desired output if the parameters were known [4]. The process noise standard deviation was chosen as $\lambda = 1.5$.

A Monte Carlo simulation procedure was conducted to compare the performance of the MAD control algorithm with the performances of the HCE and DUL algorithms, when each is applied to this problem.

TABLE I
SAMPLE AVERAGE COSTS AND STANDARD DEVIATIONS FOR EXAMPLE 1

Algorithm	OPT	HCE	DUL	MAD
Sample Mean	63.8	154	151	103
Sample standard deviation	73.2	233	251	135

TABLE II
STATISTICAL TEST FOR ALGORITHM COMPARISONS FOR EXAMPLE 1

Algorithms compared			Test statistic	Estimated improvement	Level of significance
HCE - DUL	3.63	6.67	2.54	2	3
HCE - MAD	50.94	13.6	3.72	33	0
DUL - MAD	47.21	15.	3.01	31	.13

TABLE III
SAMPLE AVERAGE COSTS AND STANDARD DEVIATIONS FOR EXAMPLE 2

Algorithm	OPT	HCE	DUL	MAD
Sample Mean	60.5	264	221	109
Sample standard deviation	73.2	440	434	137

TABLE IV
STATISTICAL TEST FOR ALGORITHM COMPARISONS FOR EXAMPLE 2

Algorithms compared			Test Statistic Δ/σ	Estimated improvement
HCE - DUL	42.96	13.69	3.13	16
HCE - MAD	154.46	29.66	5.20	59
DUL - MAD	111.50	26.9	4.14	50

Statistical tests were made on the results of 200 independent Monte Carlo runs. Each of the 200 sets of disturbances was used to generate a run for each of the three control algorithms examined. For $\Pi(0) \cdot 200 = 100$ runs, the true parameter vector was set at $\theta = \theta_1$ and for $[1 - \Pi(0)] \cdot 200 = 100$ runs it was set at $\theta = \theta_2$. Sample means and variances of the Monte Carlo costs C_i defined by (2.8) were computed.

Table I contains the results. The column labeled OPT is the performance for the same disturbances when the optimal control with θ *known* is used. This table gives the first indication of the improvement MAD gives over HCE and DUL both in mean cost reduction and reduction in the variability of the performance.

Note that Table I does not provide a rigorous argument that the actual performances (expected costs) are ordered as the sample means indicate. Appendix II presents a rigorous statistical test that provides the answer to the question of whether the expected values of the costs are different.

To carry out this test, three new data sequences are formed by taking the differences of the cost samples generated using the same random variables for each of the methods HCE, DUL, MAD. That is

$$\Delta_i^{HD} = C_i^{HCE} - C_i^{DUL} \qquad (7.5)$$

$$\Delta_i^{HM} = C_i^{HCE} - C_i^{MAD} \qquad (7.6)$$

$$\Delta_i^{DM} = C_i^{DUL} - C_i^{MAD} \qquad (7.7)$$

for $i = 1, \cdots, 200$. The sample means $\bar{\Delta}$ of the differences and their standard deviations $\sigma_{\bar{\Delta}}$ for the various algorithms are given in Table II.

Assuming that a hypothesis can be accepted only if the probability of error (level of significance) α is less than 5 percent, i.e., the confidence $(1 - \alpha)$ is at least 95 percent, the threshold against which we compare the test statistic $\bar{\Delta}/\sigma_{\bar{\Delta}}$ is $\mu = 1.65$. The test statistic has to exceed the threshold in order to accept the hypothesis. The conclusions that can be drawn for this problem from Table II are the following.

1) The hypothesis that DUL is better than HCE cannot be accepted. The estimated improvement of 2 percent is not statistically significant ($\alpha = 30$ percent is too large a probability of error to accept that DUL is better than HCE).

2) The hypothesis that MAD is better than HCE is accepted (actually with 99.99 percent confidence). The estimated improvement (decrease in cost) of 33 percent is statistically significant.

3) The hypothesis that MAD is better than DUL is accepted (actually

with 99.87 percent confidence). The estimated improvement of 31 percent is statistically significant.

Note that MAD has gone about 55 percent of the way between DUL and OPT; the latter is, however, an unachievable lower bound because it assumes the parameters *known*. The Bayesian optimal controller for *unknown* parameters (obtained from the stochastic dynamic programming) is somewhere between OPT and MAD. Thus, MAD seems to have gone "most of the way" towards the Bayesian optimum.

Example 2: This example is the same as the first one except for the cost weightings, which are

$$q = [0 \quad 1 \quad 1 \quad 1 \quad 5 \quad 50]. \qquad (7.8)$$

and the reference trajectory

$$y_r = [0.1 \quad 0.5 \quad 1 \quad 2 \quad 0.1 \quad 10]. \qquad (7.9)$$

The resulting average cost and standard deviations from 200 Monte Carlo runs are shown in Table III.

Table IV indicates the following.

1) The hypothesis that DUL is better than HCE is accepted. The estimated improvement of 16 percent is statistically significant ($\alpha < 0.1$ percent).

2) The hypotheses that MAD is better than both HCE and DUL are accepted ($\alpha < 0.001$ percent).

Also note that MAD reduces by 50 percent the cost incurred with DUL, based on the 200 Monte Carlo runs.

Next, the learning properties of the above algorithms are illustrated by presenting further results from the simulations of Example 2. Table V shows in the first part the evolution in time of the posterior probability that $\theta = \theta_1$ (averaged over 100 runs) when the true system had $\theta = \theta_1$. These probabilities all tend to unity but the active learning feature of MAD causes its probability to converge faster. Thus, active probing, the need for which is realized only by MAD, pays off. The second part of this table presents the corresponding results for the case $\theta = \theta_2$ true, where convergence to zero (as required) is again faster for MAD.

The need for active learning as sensed by MAD is illustrated in Table VI. For various possible values of the control at period 1, the MAD algorithm evaluates the future learning opportunities. For $u(1) = 4.3$, the preposterior density characterized by Π_1 and Π_2 indicates that not enough learning will take place: the contribution of J_{21} (which is the

TABLE V
EVOLUTION OF $P\{\theta = \theta_1 | Z(t)\}$

TABLE VI
COST BREAKDOWN AND LEARNING FOR MAD

$u(1)$	$\hat{J}(1)$	$\pi_1(2)$	$\pi_2(2)$	$\hat{J}_{11}(2)$	$\hat{J}_{12}(2)$	$\hat{J}_{21}(2)$	$\hat{J}_{22}(2)$
4.300	127.4	0.9398	0.05919	64.07	67.92	2090	66.39
4.500	112.0	0.9473	0.05185	64.78	67.38	1787	65.87
4.700	81.96	0.9540	0.04518	65.65	67.10	714.2	65.39
4.900	69.74	0.9602	0.03917	66.68	66.96	181.7	64.94
MAD 5.097	68.65	0.9655	0.03388	67.85	66.92	115.0	64.55
5.300	68.97	0.9704	0.02910	69.22	66.98	112.0	64.18
5.500	69.53	0.9745	0.02502	70.73	67.13	118.3	63.86
5.700	70.20	0.9781	0.02154	72.40	67.37	126.4	63.58
5.900	70.97	0.9911	0.01859	74.23	67.72	135.1	63.34

result of a mismatched controller that does not learn fast enough what the true system is) to the cost makes $\bar{J}$ [see (5.1)] large. For larger $u(1)$, the learning is faster but after a point its price exceeds the benefit.

Examination of Table VI also gives valuable insight into the problem of determining when there is value in using an actively adaptive controller like MAD: when the penalty for mismatched controllers is large and the contribution to the cost is significant.

VIII. SUMMARY AND CONCLUSIONS

The concept of preposterior analysis has been successfully used to derive an approximation to the stochastic dynamic programming equation for the control of systems with discrete-valued random parameters. The resulting algorithm, called model adaptive dual control, is the only actively adaptive controller for this class of systems. A rigorous methodology for comparison of control algorithms has been presented and used to show that the new actively adaptive controller yields statistically significant performance improvement over two state-of-the art passively adaptive controllers. The question of when it is worthwhile to use an actively adaptive controller (which is relatively expensive) versus a passively adaptive one has been also addressed. While Monte Carlo studies combined with the appropriate statistical analysis techniques are the best tool, a decomposition of the cost-to-go can be utilized to assess inexpensively whether one can expect a significant improvement when using this actively adaptive control versus a passive one. Based on our experience, the class of problems in which one can expect benefit from using an actively adaptive control is where there is heavy terminal state penalty and the control period is relatively short, i.e., passive learning does not suffice and there is opportunity and need for active learning. In general, active adaptation can be expected to improve the transient behavior in adaptive control by speeding up the adaptation process.

APPENDIX I
MOMENT MATCHING FOR THE APPROXIMATE PREPOSTERIOR DENSITY

The moment matching technique used to obtain $\Pi_l(t+1)$, $l=1,2$, in the approximate preposterior density (3.14) is now described. First consider finding the true moments $E[\Pi(t+1)|Z(t),u(t)]$, $E[\Pi^2(t+1)|Z(t),u(t)] \triangleq \overline{\Pi^2}(t+1)$. From the fundamental theorem of expectation and (3.5)

$$E[\Pi(t+1)|Z(t),u(t)] = \Pi(t) \quad (A.1)$$

$\overline{\Pi^2}(t+1)$ must be obtained by numerical integration using either (3.11) or (3.5) combined with $p[y(t+1)|Z(t),u(t)]$. The latter approach lends itself to a particularly simple and accurate integration procedure. Thus, take

$$\overline{\Pi^2}(t+1) = \int_{-\infty}^{\infty} \Pi^2[y(t+1)]p[y(t+1)|Z(t),u(t)]dy(t+1). \quad (A.2)$$

Now note from (3.5) that

$$\Pi(t+1) = \left\{ 1 + \frac{1-\Pi(t)}{\Pi(t)} \exp\left\{ -\frac{1}{2\lambda^2}\left[2\varsigma(t+1) \right.\right.\right.$$
$$\left.\left.\left. +(-1)^{j-1}\xi^2(t+1)\right]\right\}\right\}^{-1} \quad j=1 \text{ or } 2 \quad (A.3)$$

where

$$\xi(t+1) = \hat{y}_1(t+1) - \hat{y}_2(t+1) \quad (A.4)$$

and

$$\varsigma(t+1) = \xi(t+1)\left[y(t+1) - \hat{y}_j(t+1) \right]. \quad (A.5)$$

Using (A.3)–(A.5) and (3.5), the integration (A.2) may be shown to reduce to

$$\overline{\Pi^2}(t+1) = \sqrt{2}\,\lambda\xi(t+1)\left\{ \Pi(t)\int_{-\infty}^{\infty} f_1(\tau_1)e^{-\tau_1^2}d\tau_1 \right.$$
$$\left. +[1-\Pi(t)]\int_{-\infty}^{\infty} f_2(\tau_2)e^{-\tau_2^2}d\tau_2 \right\} \quad (A.3)$$

where

$$\tau_j = \frac{\varsigma(t+1)}{\sqrt{2}\,\lambda\xi(t+1)} \quad (A.4)$$

and

$$f_j(\tau_j) = \left\{ 1 + \frac{1-\Pi(t)}{\Pi(t)} \exp\left\{ -\frac{\xi(t+1)}{2\lambda^2}\right.\right.$$
$$\left.\left. \cdot\left[2\sqrt{2}\,\lambda\tau_j + (-1)^{j-1}\xi(t+1)\right]\right\}\right\}^{-2} \quad j=1,2. \quad (A.5)$$

The integrals in (A.3) reduce to simple finite length summations through the use of Hermitian quadrature [12]; this technique is described by

$$\int_{-\infty}^{\infty} f(\tau)e^{-\tau^2}d\tau = \sum_{i=1}^{I_q} H_i f(\chi_i) \quad (A.6)$$

where the χ_i are the zeros of Hermite orthogonal polynomials and the H_i are the respective Hermite coefficients. The H_i, χ_i are well tabulated [12]. The number of terms I_q in the expansion is chosen large enough to achieve desired accuracy in (A.6).

Using (A.6), (A.3) becomes

$$\overline{\Pi^2}(t+1) = \sqrt{2}\,\lambda\xi(t+1) \sum_{i=1}^{I_q} H_i\{\Pi(t)f_1(\chi_i) + [1-\Pi(t)]f_2(\chi_i)\}. \quad (A.7)$$

Equating (A.1) and (A.7) to the respective moments produced by (3.14) gives

$$\Pi(t) = \Pi(t)\Pi_1(t+1) + [1-\Pi(t)]\Pi_2(t+1) \quad (A.8)$$

$$\overline{\Pi^2}(t+1) = \Pi(t)\Pi_1^2(t+1) + [1-\Pi(t)]\Pi_2^2(t+1). \quad (A.9)$$

Using inequality (3.15), (A.8) and (A.9) then yield the desired delta function locations

$$\Pi_1(t+1) = \Pi(t) + \left\{ [1 - \Pi(t)] \left[\frac{\overline{\Pi^2}(t+1)}{\Pi(t)} - \Pi(t) \right] \right\}^{1/2} \quad \text{(A.10)}$$

$$\Pi_2(t+1) = \frac{\Pi(t)}{1 - \Pi(t)} [1 - \Pi_1(t+1)]. \quad \text{(A.11)}$$

APPENDIX II
STATISTICAL SIGNIFICANCE IN THE COMPARISON OF CONTROLLER PERFORMANCES

Suppose that a Monte Carlo simulation is performed to compare two control algorithms. The corresponding expected costs are $J^{(1)}$ and $J^{(2)}$. If S independent runs are made with the first algorithm, this yields S independent samples $C_i^{(1)}$ from a distribution with the true but unknown mean $J^{(1)}$. If the same random variables that entered into the Monte Carlo runs with the first algorithm are used to generate S runs with the second algorithm, this yields S samples $C_i^{(2)}$ from a distribution with the also unknown mean $J^{(2)}$.

The sample means

$$\bar{C}^{(j)} = \frac{1}{S} \sum_{i=1}^{S} C_i^{(j)} \quad j = 1, 2 \quad \text{(B.1)}$$

are estimates of the corresponding performances (true means). A statement that

$$\bar{C}^{(1)} < \bar{C}^{(2)} \quad \text{(B.2)}$$

implies that algorithm 1 is better than 2 *must* be qualified by a probability α of error type I [16].

Thus, the statistical test needed is

$$H_0: \quad \Delta = J^{(2)} - J^{(1)} < 0 \quad \text{(algorithm 1 not best)} \quad \text{(B.3)}$$

$$H_1: \quad \Delta = J^{(2)} - J^{(1)} > 0 \quad \text{(algorithm 1 best)}. \quad \text{(B.4)}$$

The probability of error (also called level of significance) α is defined as

$$\alpha = P\{ \text{accept } H_1 | H_0 \text{ true} \}. \quad \text{(B.5)}$$

Then, since accepting H_1 means rejecting H_0, the lower α is the less "significant" H_0 is. Thus, when we accept H_1 with a small α we are more confident in H_1 being true.

The test is carried out by examining the set of independent samples

$$\Delta_i = C_i^{(2)} - C_i^{(1)} \quad \text{(B.6)}$$

as to whether their true mean Δ can be accepted as being positive with high confidence (low α). Assuming S large enough, the hypothesis H_1 is accepted if

$$\bar{\Delta} > \mu \sigma_{\bar{\Delta}} \quad \text{(B.7)}$$

where

$$\bar{\Delta} = \frac{1}{S} \sum_{i=1}^{S} \Delta_i \quad \text{(B.8)}$$

$$\sigma_{\bar{\Delta}}^2 = \frac{1}{S^2} \sum_{i=1}^{S} (\Delta_i - \bar{\Delta})^2 \quad \text{(B.9)}$$

and, in view of the central limit theorem, μ is taken from the normal distribution tables. For example, for $\mu = 1.65$, $\alpha = 5$ percent, and for $\mu = 2.33, \alpha = 1$ percent. The corresponding confidence in the statement that algorithm 1 is superior to 2 is then $1 - \alpha$.

REFERENCES

[1] M. Aoki, *Optimization of Stochastic Systems*. New York: Academic, 1967.
[2] M. Athans *et al.*, "The stochastic control of the F-8C aircraft using a multiple model adaptive control (MMAC) method—Part I: Equilibrium flight," *IEEE Trans. Automat. Control*, vol. AC-22, pp. 768–780, Oct. 1977.
[3] Mini-Issue on F-8 DFBW Aircraft Control, *IEEE Trans. Automat. Control*, vol. AC-22, pp. 752–806, Oct. 1977
[4] K. J. Astrom, *Introduction to Stochastic Control Theory*. New York: Academic, 1970.
[5] M. Avriel, *Nonlinear Programming: Analysis and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1976
[6] Y. Bar-Shalom and E. Tse, "Concepts and Methods in Stochastic Control," in *Control and Dynamic Systems: Advances in Theory and Applications*, vol. 12, C. T. Leondes, Ed. New York: Academic, 1976.
[7] —, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Trans. Automat. Control*, vol. AC-19, pp. 494–500, Oct. 1974.
[8] Y. Bar-Shalom, R. E. Larson, and M. A. Grossberg, "Application of stochastic control theory to resource allocation under uncertainty," *IEEE Trans. Automat. Control*, vol. AC-19, pp. 1–7, Feb. 1974.
[9] Y. Bar-Shalom, "Performance adaptive resource allocation in a time-varying environment," in *Proc. IEEE Conf. Decision and Control*, 1976.
[10] —, "Effects of uncertainties on the control performance of linear systems with unknown parameters and trajectory confidence tubes," in *Annals of Economic and Social Measurements*, pp. 599–611, 1978.
[11] R. Bellman, *Adaptive Control Process: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.
[12] *CRC Handbook of Tables for Mathematics*, The Chemical Rubber Co., Cleveland, OH, 1970.
[13] J. G. Deshpande, T. N. Upadhyay, and D. G. Lainiotis, "Adaptive control of linear stochastic systems," *Automatica*, vol. 9, pp. 107–115, 1973.
[14] J. B. Farison, R. E. Graham, and R. C. Shelton, "Identification and control of linear discrete-time systems," *IEEE Trans. Automat. Control*, vol. AC-12, pp. 438–442, 1967.
[15] A. A. Feldbaum, *Optimal Control Systems*. New York: Academic, 1965.
[16] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*. New York: Macmillan, 1970.
[17] J. L. Melsa and A. P. Sage, *An Introduction to Probability and Stochastic Process*. Englewood Cliffs, NJ: Prentice-Hall, 1973, p. 25.
[18] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. Cambridge, MA: M.I.T. Press, 1972.
[19] G. N. Saridis, *Self-Organizing Control of Stochastic Systems*. New York: Dekker, 1977.
[20] G. N. Saridis and R. N. Lobbia, "Parameter identification and control of linear discrete-time systems," *IEEE Trans. Automat. Control*, vol. AC-17, pp. 52–60, 1972.
[21] G. N. Saridis and T. K. Dao, "A learning approach to the parameter adaptive self-organizing control problem," *Automatica*, vol. 9, pp. 589–598, 1972.
[22] H. A. Spang, "Optimum control of an unknown linear plant using Bayesian estimation of the error," *IEEE Trans. Automat. Control*, vol. AC-10, pp. 80–83, Feb. 1965.
[23] G. Stein and G. N. Saridis, "A parameter adaptive control technique," *Automatica*, vol. 5, pp. 731–739, 1969.
[24] E. Tse and Y. Bar-Shalom, "An actively adaptive control for discrete-time systems with random parameters," *IEEE Trans. Automat. Control*, vol. AC-18, pp. 109–117, Apr. 1973.
[25] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

Stochastic Dynamic Programming: Caution and Probing

YAAKOV BAR-SHALOM, SENIOR MEMBER, IEEE

Abstract — The purpose of this paper is to unify the concepts of caution and probing put forth by Feldbaum [14] with the mathematical technique of stochastic dynamic programming originated by Bellman [5]. The decomposition of the expected cost in a stochastic control problem, recently developed in [8], is used to assess quantitatively the caution and probing effects of the system uncertainties on the control. It is shown how in some problems, because of the uncertainties, the control becomes cautious (less aggressive) while in other problems it will probe (by becoming more aggressive) in order to enhance the estimation/identification while controlling the system. Following this a classification of stochastic control problems according to the dominant effect is discussed. This is then used to point out which are the stochastic control problems where substantial improvements can be expected from using a sophisticated algorithm versus a simple one.

I. Introduction

THIS PAPER reviews recent work in the area of stochastic control and shows how the concepts of caution and probing, originated by Feldbaum [14], can be unified with Bellman's dynamic programming technique [5], [6]. The concepts of caution and probing, developed by Feldbaum [14] about 20 years ago and also discussed in [16], dealt from an intuitive point of view with some phenomena peculiar to stochastic control problems or decision under uncertainty.

In the presence of uncertainty, modeled by random variables or stochastic processes, there is usually a deterioration of the system performance, which can be measured by an increase in the (expected) loss function compared to the deterministic case. In order to reduce the increase in the loss function the controller will tend to be "cautious," a property known in the decision theory literature as "risk aversion" [12]. This phenomenon occurs for convex loss functions that the decision maker (controller) wants to minimize, like in most control problems. On the other hand, in multistage problems where observations are made on the system at each stage, the controller might be able to carry out what has been called "active information gathering" or "probing" of the system for estimation enhancement. This is possible when the controller affects not only the state of the system but also the quality of the estimation process, i.e., has the so-called "dual effect."

This paper intends to provide a tutorial on these aspects of stochastic control by a suitable presentation of the basic concepts embodied in the stochastic dynamic programming. When the caution and probing phenomena are present in the multistage problems, the optimal solution is not known. In view of this, the insight is provided by considering a suboptimal algorithm that has the features of the optimal one.

Section II discusses the information state in the multistage control problem of a stochastic system. The formulation of the principle of optimality for stochastic systems and the resulting stochastic dynamic programming equation for additive cost functions are discussed in Section III. It is pointed out how the "preposterior analysis" technique is a direct consequence of the principle of optimality. The definition of the dual effect and the types of approximate solutions of the stochastic dynamic programming are the topic of Section IV. The "closed-loop" approximation of the stochastic dynamic programming using the "wide-sense" information state [8], [29], [30] is shown in Section V to lead to a decomposition of the expected cost into three terms. Two of these terms can be associated directly with the caution and probing effects discussed earlier giving thus a quantitative measure of these effects. It is shown in Section VI how one can classify stochastic control problems according to the dominant term in the cost decomposition. This is then illustrated via a number of examples where stochastic control problems that are probing-dominated, caution-dominated, and essentially deterministic are presented. The effect of various state weightings in the cost function and the anticipated future learning are also discussed. Conclusions are presented in Section VII.

II. THE INFORMATION STATE IN A STOCHASTIC CONTROL PROBLEM

The principle of optimality of Bellman [5] can be stated as follows for stochastic problems: at any time, whatever the present information and past decisions, the remaining decisions must constitute an optimal policy with regard to the current information set.

In the deterministic case the information set is the state of the system. This, together with the controller's subse-

quent decisions fully determines the future evolution of the system. In the stochastic case the information set is, loosely, what the controller knows about the system. This will be discussed in more detail next.

Consider the following general stochastic control problem. The state x evolves according to the equation

$$x(k+1) = f[k, x(k), u(k), v(k)] \qquad k = 0, 1, \cdots$$
$$(2.1)$$

where u is the control and v is the process noise. The measurements are described by

$$y(k) = h[k, x(k), w(k)] \qquad k = 1, \cdots \qquad (2.2)$$

where w is the measurement noise. The information set at time k is assumed to be the past measurements and controls

$$I^k = \{Y^k, U^{k-1}\} \supset I^{k-1} \qquad (2.3)$$

where

$$Y^k = \{y(j)\}_j^k, \qquad U_i^k = \{u(j)\}_j^k, \qquad (2.4)$$

and subscript $i=0$ is omitted. The inclusion property in (2.3) points to the fact that the sequence of information as assumed here is nested — each contains its predecessor.

Since (2.3) is growing with k it is of interest when a (nongrowing) information state can replace (2.3).

Note that $x(k)$ is a state only in the deterministic context when, together with U_k^{j-1}, it fully determines $x(j)$, $\forall j > k$, i.e., $x(k)$ summarizes the past of the system. The stochastic counterpart of this is the "information state."

The information state is defined as a vector-valued variable or a function that summarizes the past (i.e., it can replace I^k) when we want to characterize (probabilistically) the future evolution of the system. This is more general than the "informative statistic" of Striebel [26] which is, roughly, what the optimal controller (for the problem under consideration) needs from the past data (2.3).

It is assumed in the sequel that all the pertinent probability densities exist. Discrete-valued random variables will have a probability density function (pdf) with Dirac delta functions at the locations of the point masses.

If both sequences of process and measurement noises are white and mutually independent, then at time k the conditional probability density function of the vector $x(k)$

$$S^k = p[x(k)|I^k]^1 \qquad (2.5)$$

is an information state. This can be seen from the following. The conditional density of $x(k+1)$ can be written using Bayes' rule

$$S^{k+1} = p[x(k+1)|I^{k+1}] = \frac{1}{c} p[y(k+1)|x(k+1), I^k, u(k)]$$
$$\cdot p[x(k+1)|I^k, u(k)] \qquad (2.6)$$

where c is a normalization constant.

[1] Rigorously, the conditional density should be written $p[\cdot|Y^k, U^{k-1}]$ because this is conditioned on the sigma-algebra generated by the measurements but it is not well-defined unless the values of past controls or control functions are indicated [26]. For $k=0$ this is the prior density of the state

If the measurement noise is white ($w(k+1)$ conditioned on $x(k+1)$ has to be independent of $w(j)$, $j \leq k$, i.e.. state dependent measurement noise is allowed). then

$$p\left[y(k+1)|x(k+1), I^k, u(k)\right] = p\left[y(k+1)|x(k+1)\right] \tag{2.7}$$

(the control is anyway irrelevant in the conditioning).

For an arbitrary value of the control at k one has

$$p\left[x(k+1)|I^k, u(k)\right] = \int p\left[x(k+1)|x(k), I^k, u(k)\right]$$
$$\cdot p\left[x(k)|I^k, u(k)\right] dx(k). \tag{2.8}$$

If the process noise sequence is white and independent of the measurement noises ($v(k)$ conditioned on $x(k)$ has to be independent of $v(j-1)$, $w(j)$, $j \leq k$, i.e., state dependent process noise is allowed). then

$$p\left[x(k+1)|x(k), I^k, u(k)\right] \quad p\left[x(k+1)|x(k), u(k)\right] \tag{2.9}$$

and, since

$$p\left[x(k)|I^k, u(k)\right] = p\left[x(k)|I^k\right] = S^k \tag{2.10}$$

then. inserting (2.9) and (2.10) into (2.8) it follows that

$$p\left[x(k+1)|I^k, u(k)\right] = \phi\left[k+1, S^k, u(k)\right]. \tag{2.11}$$

Now. using (2.7) and (2.11) in (2.6) one has

$$S^{k+1} = \psi\left[k+1, S^k, y(k+1), u(k)\right]. \tag{2.12}$$

i.e.. I^k is summarized by S^k. Equation (2.12) is the recursion for the information state.

From the smoothing property of expectations it also follows that. for $j > k$,

$$p\left[x(j)|I^k, U_k^{j-1}\right] \quad E\left[S^j|I^k, U_k^{j-1}\right]$$

$$\int p\left[x(j)|x(k), I^k, U_k^{j-1}\right] p\left[x(k)|I^k\right] dx(k)$$

$$\int p\left[x(j)|x(k), U_k^{j-1}\right] S^k dx(k)$$

$$\mu\left[j, S^k, U_k^{j-1}\right] \tag{2.13}$$

where the whiteness of the process noise sequence and its independence from the measurement noises has been used again

Therefore. the whiteness and mutual independence of the two noise sequences is a sufficient condition for S^k to be an information state. It should be emphasized that the whiteness is the crucial assumption. This is equivalent to the requirement that $x(k)$ be an incompletely observed Markov process. If. for example. the process noise sequence is not white it is obvious that S^k does not summarize the past data. In this case the vector x is not a state anymore and it has to be augmented (see. e.g., [31]). This discussion points out the reason why the formulation of stochastic control problems is done with white noise sequences.

III. FROM THE PRINCIPLE OF OPTIMALITY TO STOCHASTIC DYNAMIC PROGRAMMING

Consider the problem where the number N of time steps is finite and deterministic. In general, the terminal time can be a random variable. possibly depending on the state or a decision variable. The present discussion is limited to the fixed terminal time problems. See. e.g.. [11], [18] for discussions on the free end-time problem. Denote the (scalar) cost function of the problem as

$$C = C(X^N, U^{N-1}). \tag{3.1}$$

Since this is a random variable. the minimization (in general. extremization) is done in the Bayesian approach on the expected cost

$$J = E\{C\}. \tag{3.2}$$

We assume here that the minimum and. therefore. an optimal solution (policy) exist. Otherwise. the infimum of (3.2) is to be obtained and then only an ϵ-optimal policy exists (see. e.g.. [11, p. 42]). Other approaches. like min - max and worst distribution. are also used sometimes but they are usually more difficult.

In order for (3.2) to be a well-defined criterion. the expectation must exist. i.e.. all the variables entering into the cost must be either deterministic or random (with suitable moment conditions that guarantee the existence of the expected cost). No "unknown constants" can be used in formulating stochastic control problems with the Bayesian approach.

If there are unknown system parameters. they have to be modeled as random variables with a $priori$ pdf. If these parameters are time invariant. then one has a single realization from the prior pdf. i.e.. an unknown system model generated by a probabilistic mechanism before the start of the process. In this case the minimization of the expected cost implies that we want to find the optimal policy

1) over all possible initial conditions (as specified by their pdf);

2) over all possible values of the unknown parameters (whose realization is according to the corresponding pdf) the ensemble of systems perceived by the controller in view of its uncertainty:

3) over all possible disturbance sequences.

When there are unknown time-invariant or slowly varying system parameters the stochastic controller can then be adaptive. i.e.. it will (hopefully) "learn" the system parameters during the control period.

The causality condition is that any decision function must depend only on the information set available at the time it has to be computed. i.e..

$$u(k) \quad u(k, I^k) \quad k \quad 0, 1, \cdots, N-1. \tag{3.3}$$

Since the principle of optimality states that every end part of the decision process must be optimal. the multistage optimization has to be started from the last stage. The last decision. $u(N-1)$. must be optimal with regard to the information set available when it has to be computed. i.e.. it will be obtained from the functional minimization

$$\min_{u(N-1)} E(C|I^{N-1}) \qquad (3.4)$$

where C is the cost for the entire problem.

The next to the last decision, $u(N-2)$

1) must be optimal with respect to (w.r.t.) I^{N-2} and

2) is to be made knowing that the remaining decision $u(N-1)$ will be optimal w.r.t. $I^{N-1} \supset I^{N-2}$.

Thus, the (functional) minimization that yields the decision function at $N-2$ is

$$\min_{u(N-2)} E\left[\min_{u(N-1)} E(C|I^{N-1})|I^{N-2} \right] \qquad (3.5)$$

and it uses the result of the functional minimization (3.4).

Note that the outside averaging in (3.5) is over $y(N-1)$ using the conditional density

$$p\left[y(N-1)|I^{N-2}, u(N-2) \right] \qquad (3.6)$$

parameterized by the control at $N-2$. Since this measurement is not yet available when $u(N-2)$ is to be computed but it will be available for $u(N-1)$ it is "averaged out" in (3.5).

The above-described last two steps are entirely similar to the "preposterior analysis" technique from the operations research literature discussed, e.g., in [22]. This technique is usually formulated in the following context. The first decision [here $u(N-2)$] is for information gathering by an experiment from which a posterior information will result [here $y(N-1)$] that will be used to make the last decision [here $u(N-1)$]. The prior (to the experiment) probability density of the (posterior) result of the experiment is called the "preposterior density" and in the present problem this is (3.6). Thus, one can say that preposterior analysis, which is "anticipation" (in a statistical sense, i.e., causal) of future information is a consequence of the principle of optimality.

From the above discussion it can be seen that the principle of optimality's statement that, at every stage, "the remaining decisions must constitute an optimal policy with regard to the current information set" implies the following: every decision has to use the available "hard" information (2.3) and "soft" information (3.6) about the subsequent hard information. This can be paraphrased as the optimal controller has to know how to use what it knows as well as what it knows about what it shall know.

The extension of (3.5) to the full N-stage process yields the optimal expected cost starting from the initial time as

$$J^*(0, I^0)$$

$$\min_{u(0)} E\left\{ \cdots \min_{u(N-2)} E\left[\min_{u(N-1)} E(C|I^{N-1})|I^{N-2} \right] \cdots |I^0 \right\} \qquad (3.7)$$

where I^0 is the initial information. Note that this equation does not assume any particular form for the cost function C.

For the additive cost given by

$$C(k) = c[N, x(N)] + \sum_{j=k}^{N-1} c[j, x(j), u(j)] \qquad (3.8)$$

the minimization (3.7) of $C(0)$, the cost starting from the initial time 0 yields the discrete-time stochastic dynamic programming equation. Dynamic programming can be applied only to the so-called class of "decomposable" cost functions, as pointed out in [21], [23]. The additive cost (3.8) belongs to this class.

Since

$$C^i \triangleq \sum_{j=0}^{i} c[j, x(j), u(j)] \qquad (3.9)$$

is independent of U_{i+1}^{N-1} and using the smoothing property of the expectation operator, i.e.,

$$E\left[E(\cdot|I^j)|I^k \right] = E[\cdot|I^k] \qquad \forall j > k \qquad (3.10)$$

one has from (3.7)

$$J^*(0, I^0)$$

$$= \min_{u(0)} E\left\{ \cdots \min_{u(N-2)} E\left[\min_{u(N-1)} E\left(c(N) \right. \right. \right.$$

$$\left. \left. \left. + \sum_{j=0}^{N-1} c(j)|I^{N-1} \right]|I^{N-2} \right] \cdots |I^0 \right\}$$

$$= \min_{u(0)} E\left\{ \cdots \min_{u(N-2)} E\left[C^{N-2} + \min_{u(N-1)} E\left[c(N) \right. \right. \right.$$

$$\left. \left. \left. + c(N-1)|I^{N-1} \right]|I^{N-2} \right] \cdots |I^0 \right\}$$

$$= \min_{u(0)} E\left\{ c(0) + \min_{u(1)} E\left\{ c(1) + \cdots \min_{u(N-2)} E\left[c(N-2) \right. \right. \right.$$

$$\left. \left. + \min_{u(N-1)} E\left[c(N-1) + c(N)|I^{N-1} \right]|I^{N-2} \right] \cdots |I^1 \right\}|I^0 \right\}.$$

$$(3.11)$$

In the above the cost summands have been moved to the left outside the minimizations that are not relevant for them.

Rewriting (3.11) in (backward) recursive form yields the Bellman equation

$$J^*(k, I^k) = \min_{u(k)} E\left\{ c[k, x(k), u(k)] \right.$$

$$\left. + J^*(k+1, I^{k+1})|I^k \right\} \qquad k = N-1, \cdots, 0 \qquad (3.12)$$

where $J^*(k, I^k)$ is the optimal cost-to-go from time k to the end and its dependence on the available information set at k is explicitly pointed out. The terminal condition for (3.12) is

$$J^*(N, I^N) = E\left\{ c[N, x(N)]|I^N \right\} \qquad (3.13)$$

where the last measurement is irrelevant since it is averaged out immediately.

The stochastic dynamic programming functional equation (3.12) resulted from the use of the principle of optimality embodied in (3.7) for the additive cost (3.8). The

recursion was obtained by moving to the left in (3.11) the cost summands.

An equivalent approach, based on the "basic lemma of stochastic control" [2] is as follows. This basic lemma states that

$$\min_u E[c(x,u)] = \min_u E\{E[c(x,u)|y]\}$$

$$\geqslant E \min_{u(\cdot)} E[c(x,u)|y]. \quad (3.14)$$

i.e., if a measurement y related to x is available then the minimization of the conditional expectation [the right-hand side (RHS) of (3.14)] yields the absolute minimum. This is equivalent to the statement that to minimize an integral [the outside expectation in (3.14)] is best done by minimizing the integrand at each point via the function $u(y)$, i.e., "feedback," instead of a single value for the entire integral, i.e., "open loop." In other words, moving a minimization inside a sequence of expectations, to be in front of a conditional expectation (conditioned on all the available information) is what is needed for the global minimum. Thus, based upon (3.14) the expected cost is minimized as follows:

$$\min_{u(0), \dots u(N-2), u(N-1)} E\{C|I^0\}$$

$$= \min_{u(0), \dots u(N-2), u(N-1)} E\{\cdots E[E(C|I^N)|I^{N-1}]\cdots|I^0\}$$

$$= \min_{u(0)} E\{\cdots \min_{u(N-2)} E[\min_{u(N-1)} E(C|I^{N-1})|I^{N-2}]\cdots|I^0\}.$$

$$(3.15)$$

i.e., exactly (3.7). Note that the nestedness property (2.3) of the sequence I^k was used above.

IV. DUAL EFFECT: CAUTION AND PROBING

The solution of multistage stochastic decision processes, either in the general form (3.7) or in the stochastic dynamic programming form (3.12) for an additive cost is a formidable problem. Unless an explicit form is found for the optimal cost-to-go in (3.12) one cannot solve this functional equation except numerically. The curse of dimensionality [6] afflicted upon the deterministic dynamic programming is further compounded by the expectation operators in the stochastic case making it unsolvable with a few exceptions (in addition to numerical minimization, numerical calculation of the conditional expectations also has to be carried out, which is practically impossible).

The few exceptions are the linear-quadratic problem [1], [2], [7], the linear-exponential-quadratic-Gaussian problem [24] and a linear system with a special form cost (even powers of the state up to sixth) [25].

Since one cannot obtain the optimal stochastic controller it is of interest to find suitable approximations for the stochastic dynamic programming. Such an approximation should preserve the preposterior analysis property of the principle of optimality mentioned in the previous section

and allow an assessment of the effect of uncertainties (imperfect information: present and future) on the controller and its performance.

The approximations of the stochastic dynamic programming fall in the following two classes.

1) Feedback Type Algorithms: In this case the control depends only on the current information

$$u(k) = u(k, I^k) \quad (4.1)$$

but does not use the prior statistical description of the future posterior information

$$p[y(j+1)|I^j], \quad j \geqslant k. \quad (4.2)$$

2) Closed-Loop Type Algorithms: Such a controller utilizes feedback (4.1) *and* anticipates future feedback via (4.2), i.e., that the loop will stay closed.

Feldbaum [14] introduced the concept of dual effect in the control of stochastic dynamic systems. In a stochastic problem the control has, in general, two effects.

1) It affects the state (control action).

2) It affects the uncertainty of the state (augmented by the possibly unknown parameters).

A rather general mathematical definition of this has been given in [7] in terms of conditional central moments of the state vector. To illustrate it, let the conditional covariance of the state at k be

$$\Sigma(k|k) = E\{[x(k) - \hat{x}(k|k)][x(k) - \hat{x}(k|k)]'|I^k\}$$

$$(4.3)$$

where $\hat{x}(k|k)$ denotes the conditional mean. Then if $\Sigma(k|k)$ does not depend on the past controls U^{k-1}, the control has no dual effect (of second order), i.e., it is neutral. This is the case in linear dynamic systems with additive but not necessarily Gaussian noise [7], [32]. In nonlinear systems the state estimation accuracy is in general control dependent—the control has a dual effect.

If the system has unknown parameters, modeled as a realization of a vector valued random variable, the control values will affect, in general, the information about them derived from the measurements. Since having more accurate estimates of the system parameters is intuitively beneficial for the controller, the idea that the controller should enhance their identification is appealing. The initial control should account for the fact that it is applied to a system with parameters drawn from the prior distribution *and* for the fact that their value can be further identified during the process. This is the adaptive or learning feature of the controller. A simple example that illustrates the dual effect of the control is given in the Appendix.

Therefore, the controller can be used for "active information storage" (estimation enhancement or uncertainty reduction) via what has been called probing [14]. Note that only a "closed-loop" algorithm can do this active information gathering. On the other hand, the existence of uncertainty in the system, might have another effect. Since, in general, uncertainty in the system will increase the expected cost, the controller should be "cautious" not to

increase further the effect of the existing uncertainties on the cost. A simple example to illustrate this "caution" effect is also given in the Appendix.

The open-loop feedback (OLF) control [1], which belongs to the feedback class, works well in some problems. Nevertheless, it can suffer from the "turn-off" phenomenon which can be avoided only by a closed-loop controller [15], [36]. As pointed out in [7] the optimal solution of the linear-quadratic control problem belongs to the feedback class because in this problem the control has no dual effect. Among the algorithms that belong to the feedback class are the heuristic certainty equivalence ("enforced separation") [10], [28], the self-tuning regulator [3], the cautious control [36], and the multiple model partitioned control [4], [13]. Algorithms of the closed-loop type are the wide-sense adaptive [8], [29], [30], the dual controllers of [27], [36], the innovations dual controller of [20], and the model adaptive dual controller for multiple models [37].

V. CAUTION AND PROBING EFFECTS FROM THE STOCHASTIC DYNAMIC PROGRAMMING

The previous discussion pointed out qualitatively that a controller

1) has a direct control effect on the state;

2) can perform active information gathering (probing) to improve the accuracy of subsequent control actions; and

3) has to be cautious because of the existing uncertainties in the system.

While there is no universal agreement on the notions of caution and probing this author believes these concepts are valuable in the derivation of suboptimal algorithms. In this section a quantification of the above properties is presented. This is obtained by an approximation of the optimal cost from the stochastic dynamic programming that results in a decomposition of the cost into three terms, each associated with one of the above items.

The stochastic dynamic programming equation (3.12) is approximated as follows [8], [29], [30]. First, instead of the exact information state, the following approximate "wide-sense" information state is used:

$$\mathscr{P}^k = \{\hat{x}(k|k), \Sigma(k|k)\}, \tag{5.1}$$

i.e., the (approximate) conditional mean and covariance of $x(k)$ obtained, e.g., via an extended Kalman filter. The use of this "quasi-sufficient statistic" is needed for an algorithm that is implementable. Assume now that the system is at time k and a closed-loop control (in the sense defined earlier) is to be computed using $\mathscr{P}^k$ and the present knowledge (statistical) about the future observations.

The principle of optimality with the information state (5.1) yields the following stochastic dynamic programming equation for the closed-loop-optimal expected cost-to-go at time k

$$J^*(k, \mathscr{P}^k) = \min_{u(k)} E\{c[k, x(k), u(k)]$$

$$+ J^*(k+1, \mathscr{P}^{k+1})|\mathscr{P}^k\}. \tag{5.2}$$

The main problem is to obtain an approximate expression for $E\{J^*(k+1, \mathscr{P}^{k+1})|\mathscr{P}^k\}$ preserving its closed-loop feature, i.e., this expression should incorporate the "value" of the future observations. In order to find an explicit solution, the cost-to-go $C(k+1)$ defined in (3.8) is expanded about a nominal trajectory (designated by subscript 0) generated by the recursion

$$x_0(j+1) = f[j, x_0(j), u_0(j), \bar{v}(j)],$$
$$j = k+1, \cdots, N-1 \tag{5.3}$$

where $u_0(j)$, $j = k+1, \cdots, N-1$ is a sequence of nominal controls and $\bar{v}(j)$ is the mean of the process noise. The initial condition $x_0(k+1)$ is taken as the predicted value of the state at $k+1$ given $\mathscr{P}^k$ and the control (yet to be found) $u(k)$. The expansion of the cost-to-go from time $k+1$ is

$$C(k+1) = C_0(k+1) + \Delta C_0(k+1) \tag{5.4}$$

where $C_0(k+1)$ is the cost along the nominal (ignoring all the uncertainties) and $\Delta C_0(k+1)$ is the variation of the cost about the nominal with terms up to second order obtained from a Taylor expansion, which will capture the stochastic effects. The approximation of the closed-loop-optimal expected cost-to-go from time $k+1$ is done now as follows:

$$J^*(k+1) = C_0(k+1) + \Delta J_0^*(k+1) \tag{5.5}$$

where the optimal "closed-loop" perturbation cost is

$$\Delta J_0^*(k+1)$$

$$= \min_{\delta u(k+1)} E\left\{ \cdots \min_{\delta u(N-1)} E[\Delta C_0(k+1)|\mathscr{P}^{N-1}] \cdots |\mathscr{P}^{k+1}\right\} \tag{5.6}$$

and $\delta u(k) = u(k) - u_0(k)$. This minimization problem is quadratic since, by construction, $\Delta C_0(k+1)$ is quadratic in $\delta u(j)$, $k+1 \leq j \leq N-1$ as well as in the variations about the nominal trajectory, $\Delta x(j) = x(j) - x_0(j)$, $k+1 \leq j \leq N$. Using a Taylor series expansion of (2.1) and including second-order terms results in a set of perturbation state equations in $\delta x(j)$ with $\delta x(k+1) = x(k+1) - x_0(k+1)$ as an initial condition. Thus, the problem posed in (5.6) consists of minimizing a quadratic cost given a quadratic system of state equations, and is somewhat similar to the linear-quadratic control problem. Then, by assuming a solution quadratic in the perturbed state (i.e., neglecting higher order terms) and evaluating the expectations permits the optimal closed-loop (CL) cost-to-go to be obtained explicitly. See [8] for the development of the details. This result, obviously, depends on the approximations used in the derivation.

The Cost Decomposition

The explicit expression of the (approximate) cost obtained can be decomposed as follows:

$$J^{CL}(k) = J_D(k) + J_c(k) + J_P(k) \tag{5.7}$$

where the subscript D stands for deterministic, C stands for caution, and P stands for probing components.

It will be assumed, for simplicity, that

$$c[k, x(k), u(k)] = c_1[k, x(k)] + c_2[k, u(k)] \quad (5.8)$$

and that the process noise, whose covariance is V, enters additively in (2.1). Then the deterministic component of the cost-to-go is, excluding c_1 (which does not depend on the control) is given by

$$J_D(k) \doteq c_2[k, u(k)] + C_0(k+1) + \gamma_0(k+1) \quad (5.9)$$

and the stochastic terms obtained via the perturbation problem are

$$J_C(k) \doteq 1/2 \, \mathrm{tr}[K_0(k+1)\Sigma(k+1|k)]$$

$$+ 1/2 \sum_{j=k+1}^{N-1} \mathrm{tr}[K_0(j+1)V(j)] \quad (5.10)$$

$$J_P(k) \doteq 1/2 \sum_{j=k+1}^{N-1} \mathrm{tr}[d_{0,uu}(j)\Sigma_0(j|j)]. \quad (5.11)$$

Σ is the covariance of the augmented state and γ, K, and d are given by appropriate recursions detailed in [8].

The stochastic term (5.10) reflects the effect of the uncertainty at time k summarized by $\Sigma(k|k)$ and subsequent process noises on the cost. These uncertainties cannot be affected by $u(k)$ but their weightings do depend on it, e.g., $\Sigma(k+1|k)$ depends on $\Sigma(k|k)$ and $u(k)$. The effect of these uncontrollable uncertainties on the cost should be minimized by the control; this term indicates the need for the control to be cautious and thus is called *caution term*. The stochastic term (5.11) accounts for the effect of uncertainties when subsequent decisions (corrective actions) will be made. The weighting of these future uncertainties is nonnegative ($d_{0,uu}$ is positive semidefinite). If the control can reduce by probing (experimentation) the future updated covariance, it can thus reduce the cost. The weighting matrix $d_{0,uu}$ yields approximately the *value of future information* for the problem under consideration. Therefore, this is called the *probing* term. Note that even if the control has no dual effect, i.e., it does not affect the future covariance Σ of the augmented state (which includes the random parameters), the weighting of these covariances might still be affected by the control. Therefore, this (admittedly approximate) procedure accounts not only for the dual effect but all the stochastic effects in the performance index.

Thus, starting from the stochastic dynamic programming one can see the following: the benefit of probing is weighted by its cost and a compromise is chosen such as to minimize the sum of the deterministic, caution, and probing terms. The minimization of J^{CL} will also achieve a tradeoff between the present and future actions according to the information available at the time the corresponding decisions are made.

The closed-loop control $u(k)$ is found from the minimization of (5.7) using a search procedure. At every k to each control $u(k)$ for which (5.7) is evaluated during the search there corresponds a predicted state and to this predicted state a sequence of deterministic controls is attached that defines the nominal trajectory. The only use of the nominals and perturbations is to make possible the evaluation of the cost-to-go optimized in a closed-loop manner. This procedure is repeated at every time a new control is to be obtained.

The "quality" of the approximations used in the derivations outlined above, in particular, the second-order expansions, is an open question. Only extensive Monte Carlo simulations with rigorous comparison with other algorithms (see, e.g., [37]) can answer these questions. For some problems [29], [30] significant performance improvements have been found. In other cases where probing is not significant the CL algorithm performed close to the OLF [8].

The cost decomposition is believed to provide the only insight we now have towards the understanding of complex stochastic control problems for which the optimal solution is unknown. Furthermore, the classification of various stochastic control problems presented in the next section, which is based on this decomposition, can be used as a tool to assess for which nonlinear problems stochastic control algorithms can provide significant performance improvements.

VI. IMPLICATIONS OF THE COST DECOMPOSITION AND EXAMPLES

The decomposition of J^{CL} presented above yields an explicit evaluation of the tradeoffs between direct control, active probing, and a cautious action on the part of the controller. Thus, the ability of the control to affect learning as well as steer the system to its targets can be numerically evaluated using this decomposition. This is a particularly attractive feature for it captures both the need (and desire) of the controller to extract more information from the system as well as the aversion for drastic actions which may result in undesirable outcomes (risk aversion [12]). Furthermore, this also gives indication whether the uncertainty dominates the problem when the stochastic part of the cost ($J_C + J_P$) exceeds significantly the deterministic part (J_D).

If the uncertainty dominates the problem, then one can distinguish two cases.

1) The caution component J_C dominates. Then, since this is "uncontrollable" uncertainty, one has a highly uncertain model which cannot be improved in the course of the control period.

2) The probing component J_P dominates. Then, with the dual effect of the control, one can reduce the uncertainty of the model—thus, the model, while uncertain at the beginning, might prove to be ultimately adequate for the control problem under consideration.

A third case occurs when we have the following.

3) The deterministic component of the cost J_D dominates; then the parameter uncertainties are of no significant consequence.

The last case is the most desirable because then the controller can be of the certainty equivalence type [7], i.e., it can ignore the uncertainties by replacing all the random variables by their (conditional) means. This is the least expensive algorithm because it is essentially deterministic and will yield near optimum performance. However, the stochastic control approach outlined above has to be used to reach this conclusion.

Wonham [33] stated, about ten years ago, the following. In the case of (stochastic) feedback controls the general conclusion is that only marginal improvement can be obtained (over a controller ignoring the stochastic features), unless the disturbance level is very high; in this case the fractional improvement may be large but the system is useless anyway.

This statement implies that with high-level disturbances (in which one can include large parameter uncertainties) one has a "hopeless" situation. The other extreme is the situation with low level disturbances. These two situations seem to match, respectively, cases 1) and 3) from above. It was also pointed out in [33] that Feldbaum's dual control which probes the system might hold the promise of useful applications of stochastic control. However, at that time it was not clear whether there are such problems and, if yes, then how to obtain a (dual) controller that can effectively probe the system to reduce uncertainties. The wide-sense dual (or stochastic closed-loop) control algorithm [8], [29], presented in Section V, can then be used to obtain significant performance improvement.

As will be shown in the sequel, the cost decomposition presented above can answer affirmatively the question whether there are probing-dominated stochastic control problems, i.e., problems falling in case 2) from above.

In the following a number of examples are discussed to illustrate the usefulness of the cost decomposition and its implications. Some of these examples have appeared earlier in the literature and they are reexamined in light of the recently gained quantitative understanding of the caution and probing effects from the cost decomposition.

A. A Probing-Dominated Problem (Terminal Guidance)

The first example is the interception problem from [30]. In this case a third-order linear system with six unknown (random) parameters and both process and measurement noises was considered. The augmented nine-dimensional state (for which the dynamic equation is obviously nonlinear) had an initial estimate and an associated covariance. The elements of this covariance matrix corresponding to the parameters reflected the fact the initial estimates of the parameters were poor. The goal was to steer one of the (proper) state components to a target value by the terminal time, which was $N = 20$. This was expressed by a quadratic term for the terminal state. There was no cost associated
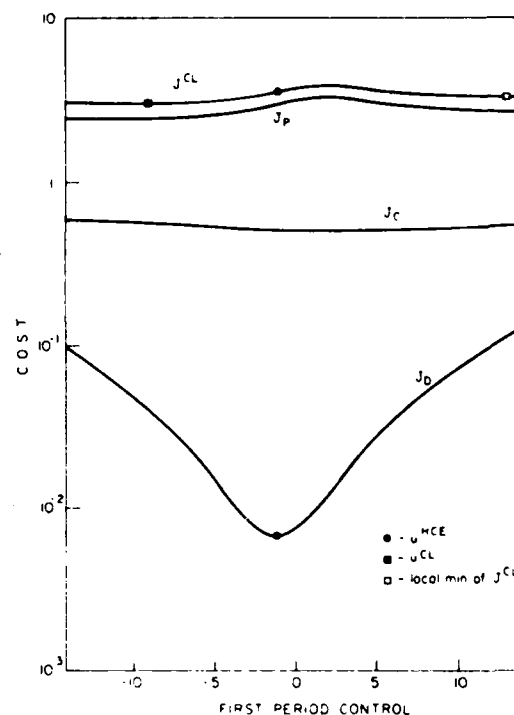

Fig. 1. Cost decomposition for a probing-dominated stochastic control problem (terminal guidance for a third-order system with six unknown parameters).

with the state prior to the terminal time and the cost weighting of the control, also entering quadratically, was low.

Fig. 1 presents the plot of the cost decomposition for the first period control. It can be seen that this is a probing-dominated stochastic control problem: the probing component of the cost is approximately 80 percent of the total cost.

The performance of the wide-sense dual [or closed-loop (CL)] control described in Section V was compared in [30] via Monte Carlo runs to the HCE (heuristic certainty equivalence) where the parameters' estimates were used as if they were the true values. The observed improvement of the CL algorithm versus HCE was, from (the modest number of) 20 Monte Carlo runs, around 85 percent [30]. This fractional improvement is quite close to the share of the probing cost from the total as indicated above. The CL controller, via its dual effect helped identify the system, i.e., it was actively adaptive and this was the key factor in its better performance. This decomposition, which was not known at the time of the original work [30], can now be used to provide the explanation for the observed performance improvement.

An important observation is that the probing component of the cost is not convex — the parameter identification is enhanced by large magnitude first period control values, both negative and positive. This lack of convexity of the probing component leads to local minima, as can be seem from Fig. 1. This phenomenon was pointed out in [27], [36]. The behavior of the multiple minima is discussed later in more detail.

The example discussed above, which is of the terminal state penalty type, belongs to the second class of problems, i.e., probing dominated.

B. A Caution-Dominated and an Essentially Deterministic Problem (Econometric Models)

Two additional problems, derived from econometrics are discussed next. Both are macroeconometric models of the U.S., derived from the same data but under different assumptions. For a concise description of the models see [9], [10]. The first econometric model has three states (gross national product, investment, and consumption), is driven by the government expenditures input, and has five unknown parameters characterized by an initial estimate and covariance matrix. The second econometric model has 11 states (as above plus increments of these variables and some lagged values), same input, and three unknown parameters.

The first model was obtained by Kendrick using ordinary least squares [17] while the second, more elaborate model, was obtained by Wall using the full information maximum likelihood method [34], [35]. The cost was quadratic in the deviations of the three economic variables and the input from target values along the entire trajectory consisting of seven periods (economic quarters).

The analysis of the cost $J^{CL}(0)$ for the first econometric model, shown in Fig. 2, points to the fact that this problem is dominated by the caution term. This is due to the relatively large uncertainties in the initial parameter estimates. The probing component is negligible — this problem is completely dominated by the initial uncertainty — it belongs to the first class defined at the beginning of the section. Note that both the caution as well as the probing term tend to reduce the value of u^{CL} versus u^{HCE}, i.e., they are not conflicting in this case.

Fig. 3 shows the cost for the second econometric model. The deterministic component dominates here and $u^{CL}(0)$ is very close to $u^{HCE}(0)$. The probing component is again negligible. This problem belongs to the third class — it is essentially deterministic.

C. A Scalar Problem: Parametric Study of the Cost Shape

Another example of the application of the cost decomposition deals with a scalar linear system over $N = 2$ time periods discussed in [19].

$$x(k+1) = ax(k) + bu(k) + v(k) \qquad k = 0,1 \quad (6.1)$$

with $a = 0.7$ known, the unknown input gain b with initial estimate $\hat{b}(0) = 0.6$, and variance $\sigma_b^2(0)$. The process noise $v(k)$ is zero mean, white with variance V. The goal is to keep the state x, which is perfectly observed, around zero. This is expressed by the quadratic cost

$$C = 1/2 Q(2)x^2(2) + 1/2 r[u^2(0) + u^2(1)] \qquad (6.2)$$

with terminal state weighting $Q(2)$ and control weighting $r = 0.1$. The initial state is $x(0) = 1$.
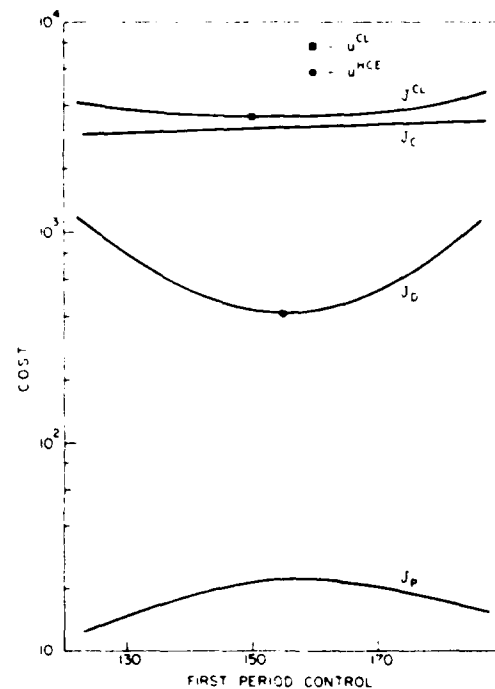

Fig. 2. Cost decomposition for a caution-dominated stochastic control problem (third-order econometric model with five unknown parameters).
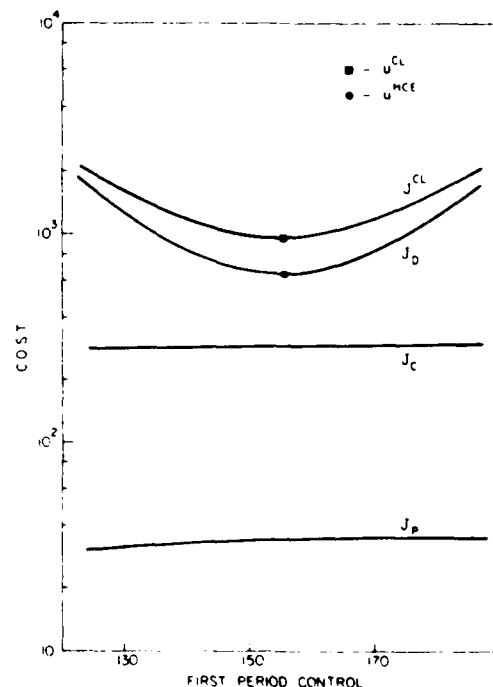

Fig. 3. Cost decomposition for an essentially deterministic stochastic control problem (11th-order econometric model with three unknown parameters).

Fig. 4 presents the cost decomposition at $k = 0$ (first period) for the initial gain uncertainty $\sigma_b^2(0) = 0.52$ and process noise variance $V = 0.2$ and terminal state weighting $Q(2) = 10$. The probing component of the cost, which varies drastically with the control, yields two minima for the total cost. It is of interest to see how these minima behave as the terminal state weighting changes. This is illustrated in Fig. 5. For even larger terminal weighting the two minima get further apart while for a lower weighting, $Q(2) = 1$, there is
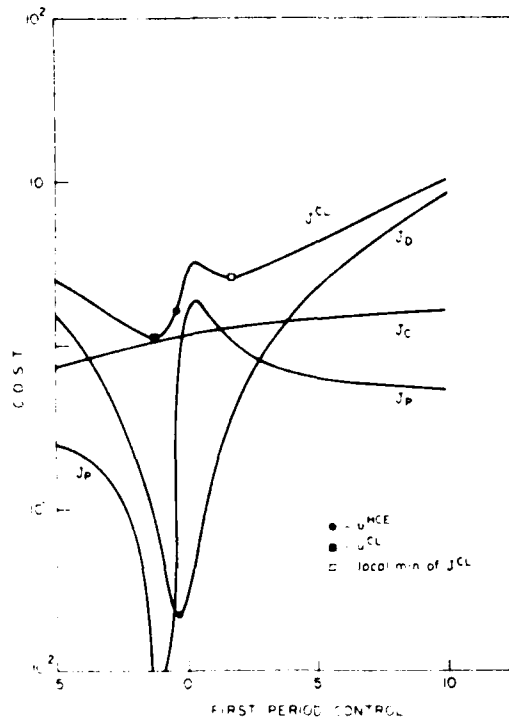
Fig 4   Cost decomposition for a two-stage problem with unknown input gain (scalar system)

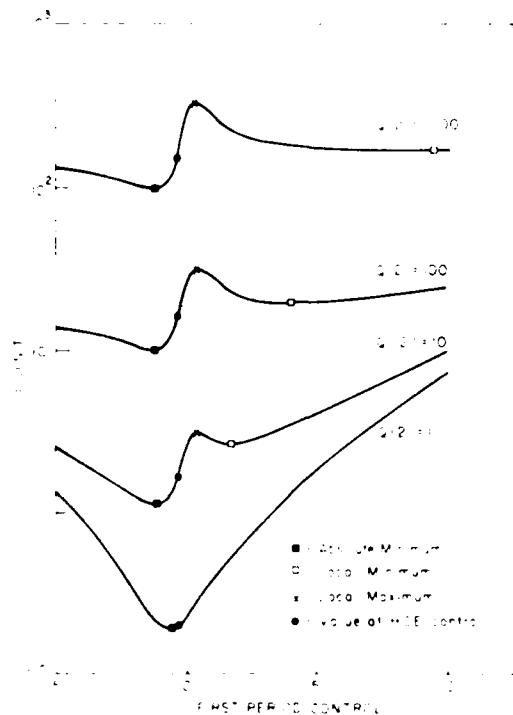Fig 6   Effect of the anticipated future learning on the control (scalar system)

Fig 5   The closed-loop cost for various terminal state weightings (scalar system)

only one minimum left. In this latter case the lighter terminal penalty does not justify a major control effort to identify accurately the parameter b and u^{CL} is quite close to u^{HCE}.

Another aspect of interest is how the anticipated future learning changes the present behavior of the CL controller. To this purpose the variance of the process noise was varied. Fig. 6 shows the cost $J^{CL}(0)$ for $Q(2) = 1000$, $\sigma_b^2 = 2$,
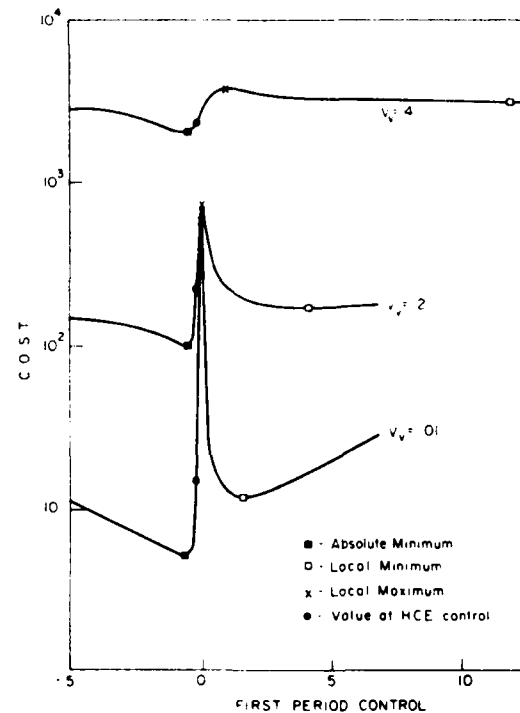
and various values of V. For large process noise variance, less learning is anticipated and the cost curve is relatively flat, even though it has two minima, wide apart. For low process noise variance the cost curve has a very high maximum at $u(0) = 0$ (when no learning of b occurs) and then two sharp minima around this point.

VII.   CONCLUSIONS

While still very few stochastic control problems have been solved optimally, insight into such problems can be gained by using the decomposition of the expected cost. This decomposition, based on the stochastic dynamic programming, yields three cost components: one deterministic and two stochastic ones. The stochastic terms quantify the effect of the various uncertainties on the performance index. The effects these stochastic terms have been associated with Feldbaum's concepts of caution and probing. Furthermore, this decomposition revealed three classes of stochastic control problems: caution dominated, probing dominated, and essentially deterministic. This, admittedly fuzzy, classification pointed out that there are stochastic control problems where significant improvements can be expected when using an appropriate sophisticated control algorithm. The examples show that one can assess, before extensive simulations, whether significant performance improvement can be expected in a stochastic control problem. It has also been shown that the various cost components can vary drastically with changes in the performance index weightings. The probing component of the cost can be nonconvex thus leading to local minima in the total cost.

APPENDIX
SIMPLE EXAMPLES OF PROBING AND CAUTION

Consider the scalar system

$$x(k+1) = ax(k) + bu(k) + v(k) \qquad (A.1)$$

with a known, b an unknown parameter with prior mean $\hat{b}(0)$ and variance $\sigma_b^2(0)$, and $v(k)$ a zero-mean white noise sequence with variance σ_v^2. Letting

$$I^k = \{ X^k, U^{k-1} \}, \qquad (A.2)$$

i.e., perfect state observations, it follows that the uncertainty about parameter b at time $k+1$ is, from a standard least-squares argument, dependent on the control at k as follows:

$$\sigma_b^2(k+1) = \frac{\sigma_b^2(k)\sigma_v^2}{\sigma_b^2(k)u^2(k) + \sigma_v^2}. \qquad (A.3)$$

This clearly illustrates the control's dual effect, in addition to its effect on the state the control also affects the future information accuracy.

Consider next the same system with the (one-step horizon or myopic) cost

$$C(k) = x^2(k+1) + \lambda u^2(k). \qquad (A.4)$$

The control that minimizes

$$J(k) = E_v^1 \{ C(k) | I^k \} \qquad (A.5)$$

can be obtained easily as

$$u^*(k) = -\frac{ax(k)\hat{b}(k)}{\hat{b}^2(k) + \sigma_b^2(k) + \lambda}. \qquad (A.6)$$

Note that, because of the myopicity of the cost (A.4), this controller ignores any possibility of learning. On the other hand, because of the uncertainty in b, this control can be very cautious — a large variance $\sigma_b^2(k)$ can decrease significantly the value of the control in (A.6) compared to the case where there is no uncertainty in b or when this uncertainty is ignored as an HCE controller would do

$$u^{HCE}(k) = -\frac{ax(k)\hat{b}(k)}{\hat{b}^2(k) + \lambda}. \qquad (A.7)$$

The optimal myopic controller (A.6) can then exhibit the turn-off phenomenon [15], [36] — it can be small because of large uncertainty in b and this will then prevent, according to (A.3), the reduction of this uncertainty.

ACKNOWLEDGMENT

REFERENCES

[1] M. Aoki, *Optimization of Stochastic Systems*. New York: Academic, 1967.

[2] K. J. Astrom, *Introduction to Stochastic Control Theory*. New York: Academic, 1976.

[3] K. J. Astrom, U. Borisson, L. Ljung, and B. Wittenmark, "Theory and application of self-tuning regulators," *Automatica*, vol. 13, pp. 457–476, Sept. 1977.

[4] M. Athans *et al.*, "The stochastic control of the F-8C aircraft using a multiple model adaptive control (MMAC) method — Part I: Equilibrium flight," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 768–780, Oct. 1977.

[5] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.

[6] —, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.

[7] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 494–500, Oct. 1974.

[8] ——, "Caution, probing, and the value of information in the control of uncertain systems," *Ann. Econ. Social Measurement*, vol. 5, pp. 323–337, Summer 1976.

[9] Y. Bar-Shalom, "Effects of uncertainties on the control performance of linear systems with unknown parameters and trajectory confidence tubes," *Ann. Econ. Social Measurement*, vol. 6, pp. 599–611, 1978.

[10] Y. Bar-Shalom and K. D. Wall, "Dual adaptive control and uncertainty effects in macroeconomic system optimization," *Automatica*, vol. 16, pp. 147–156, 1980.

[11] D. P. Bertsekas, *Dynamic Programming and Stochastic Control*. New York: Academic, 1976.

[12] M. DeGroot, *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

[13] J. G. Deshpande, T. N. Upadhyay, and D. G. Lainiotis, "Adaptive control of linear stochastic systems," *Automatica*, vol. 9, pp. 107–115, 1973.

[14] A. A. Feldbaum, *Optimal Control Systems*. New York: Academic, 1965.

[15] D. J. Hughes and O. L. R. Jacobs, "Turn-off, escape and probing in nonlinear stochastic control," in *Proc. IFAC Symp. Adaptive Contr.*, Budapest, Hungary, Sept. 1974.

[16] O. L. R. Jacobs and J. W. Patchell, "Caution and probing in stochastic control," *Int. J. Contr.*, vol. 15, pp. 189–199, 1972.

[17] D. Kendrick, "Adaptive control of macroeconomic models with measurement error," in *Optimal Control for Econometric Models*, S. Holly, B. Rustem, and M. Zarrop, Eds. London: Macmillan, 1979.

[18] H. Kushner, *Introduction to Stochastic Control*. New York: Holt, 1971.

[19] E. C. MacRae, "Linear decision with experimentation," *Ann. Econ. Social Measurement*, vol. 1, pp. 437–447, 1972.

[20] R. Milito, C. Padilla, and R. Padilla, "An innovations approach to dual control," *IEEE Trans. Automat. Contr.*, vol. AC-26, 1981.

[21] G. L. Nemhauser, *Introduction to Dynamic Programming*. New York: Wiley, 1967.

[22] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*. Cambridge, MA: M.I.T. Press, 1972.

[23] M. H. Richardson, G. W. Keller, and R. E. Larson, "Principle of optimality and dynamic programming — Forward & backward," unpublished manuscript, 1972.

[24] J. L. Speyer, J. J. Deyst, and D. H. Jacobson, "Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 358–366, Aug. 1974.

[25] J. L. Speyer, "A nonlinear control law for a stochastic infinite time problem," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 560–564, Aug. 1976.

[26] C. Striebel, "Sufficient statistics in the optimum control of stochastic systems," *J. Math. Anal. Appl.*, vol. 12, pp. 576–592, Dec. 1965.

[27] J. Sternby, "A regulator for time varying stochastic systems," in *Proc. 7th IFAC World Congr.*, Helsinki, Finland, June 1978.

[28] G. Stein and G. N. Saridis, "A parameter adaptive control technique," *Automatica*, vol. 5, pp. 731–739, 1969.

[29] E. Tse, Y. Bar-Shalom, and L. Meier, "Wide-sense adaptive dual control of stochastic nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 98–108, Apr. 1973.

[30] E. Tse and Y. Bar-Shalom, "An actively adaptive control for discrete time systems with random parameters," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 109–117, Apr. 1973.

[31] —, "Generalized certainty equivalence and dual effect in stochastic control," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 817–819, Dec. 1975.

[32] W. M. Wonham, "On the separation theorem of stochastic control," *SIAM J. Contr.*, vol. 6, no. 2, pp. 312–326, 1968.

[33] —, "Optimal stochastic control," *Automatica*, vol. 5, pp. 113–118, Jan. 1969.

[34] K. D. Wall, "FIML estimation of rational distributed lag structural form models," *Ann. Econ. Social Measurement*, vol. 5, Winter 1976.

[35] K. D. Wall, Y. Bar-Shalom, and D. Kendrick, "Adaptive control of macroeconometric models with correlated measurement errors," presented at the 5th NBER Conf. Stochastic Contr. Econ., New Haven, CT, May 1977.

[36] B. Wittenmark, "An active suboptimal dual controller for systems with stochastic parameters," *Automat. Contr. Theory Appl.*, Canada, vol. 3, pp. 13–19, 1975.

[37] C. J. Wenk and Y. Bar-Shalom, "A multiple model adaptive dual control algorithm for stochastic systems with unknown parameters," *IEEE Trans. Automat. Contr.*, vol. AC-25. pp. 703–710, Aug. 1980.

Yaakov Bar-Shalom (S'63–M'66–SM'80) was born on May 11, 1941. He received the B.S. cum laude and M.S. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1963 and 1967, respectively, and the Ph.D. degree from Princeton University, Princeton, NJ, in February 1970, all in electrical engineering.

From 1963 to 1966 he served in the Israel Defense Army as an Electronics Officer. From 1966 to 1967 he was a Teaching Assistant in the Department of Electrical Engineering of the Technion. Between 1967 and 1970 he was at Princeton University, first as a Research Assistant, then as a Research Associate. During the academic year 1968–1969 he was appointed Charles Grosvenor Osgood Fellow in recognition for outstanding performance. From 1970 to 1976 he was a Senior Research Engineer with Systems Control, Inc., Palo Alto, CA. At SCI he has developed and applied advanced stochastic control and estimation procedures to problems in the areas of defense, environmental, and economic systems. Between 1973 and 1976 he taught graduate courses in system theory at the University of Santa Clara. Currently, he is Professor of Electrical Engineering and Computer Science, University of Connecticut, Storrs. He has been consulting to Systems Control, Inc., ORI, Inc., System Development Corporation, C.S. Draper Laboratory, Bolt, Beranek and Newman, General Electric and CTEC, Inc. His research interests are in stochastic adaptive control, resource allocation under uncertainty, estimation theory, decision theory, system identification, and modeling, analysis and control of environmental and economic systems. During 1976 and 1977 he served as Chairman of the Technical Committee on Stochastic Control and Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and currently is Associate Editor of *Automatica*. He was Chairman of the 18th IEEE Symposium on Adaptive Process in 1979 and is Program Chairman of the 1982 American Control Conference. Also, he is Guest Associate Editor of this Special Issue.

Dr. Bar-Shalom is a member of Eta Kappa Nu and Sigma Xi.

STOCHASTIC CONTROL AND IDENTIFICATION ENHANCEMENT FOR THE FLUTTER SUPPRESSION PROBLEM

Y. Bar-Shalom and J. Molusis

Electrical Engineering and Computer Science Department, University of
Connecticut, Storrs, Connecticut 06268, USA

Abstract. The topic of this paper is the application of some recent results
in stochastic control to an aerospace problem where there are large
uncertainties in the dynamics of the plant to be controlled. An approxima-
tion to the stochastic Dynamic Programming is considered that results in an
adaptive control of the "closed-loop" type: it utilizes feedback (latest
state and parameter estimates and their uncertainties) as well as their
anticipated future uncertainties - it anticipates (subject to causality)
subsequent feedback. This algorithm has the feature that allows the control
to enhance the parameter identification in real time. This is done using
the control's dual effect: the control can affect the state as well as the
(augmented) state uncertainty and thus can reduce the uncertainty about some
parameters. A flight control application in which stochastic adaptive
control appears to offer significant payoff is the active control of air-
craft wing-store flutter. Improved flutter suppression can be accomplished
with an adaptive controller that has the capability to learn and identify
the flutter modes during the flight.

1. INTRODUCTION

The topic of this paper is the application of
some recent results in stochastic control to an
aerospace problem where there are large uncer-
tainties in the dynamics of the plant to be
controlled. While the stochastic Dynamic
Programming [B1,B2] yields, in principle, the
solution to general stochastic control problems,
the curse of dimensionality prevents its appli-
cation to nonlinear problems. An important
class of problems is the one of linear systems
with unknown and possibly time varying para-
meters. Such a system is nonlinear in the
augmented state, which is made up of the
proper state and the unknown parameters.

It was pointed out in [B3] that the optimal
stochastic control depends, in general, on

(*i*)   the current information (e.g., the
          latest estimate of the state and
          parameters)

(*ii*)   the quality of the current informa-
          tion (represented, e.g., by the
          covariance associated with the
          above mentioned estimates)

(*iii*)   the anticipated quality of the sub-
          sequent (future) information

The well-known optimal solution of the Linear
Quadratic Gaussian Problem (without unknown
parameters) has the so-called Certainty
Equivalence property: the resulting feedback

control has the same gain as the corres-
ponding deterministic problem and only uses
the state estimate instead of the (unavail-
able) state. This solution exhibits only
feature (*i*) from above - it is independent
of the quality of the state estimate. The
"Heuristic Certainty Equivalence" (HCE)
algorithm for linear systems with unknown
parameters consists of the following: the
parameters are estimated in real time and
the feedback gain is computed using the
latest parameter estimates as if they were
the true values [S1].*

This algorithm, while adaptive, does not
take into consideration the quality of the
parameter estimates.

An approximation to the stochastic Dynamic
Programming was presented in [T1,T2,B4].
In the terminology of [B3], the resulting
adaptive control is of the "closed-loop"(CL)
type: it utilizes feedback (latest state
and parameter estimates and their uncer-
tainties) as well as their anticipated
future uncertainties - it anticipates (sub-
ject to causality) subsequent feedback.
This algorithm has all three features (*i*)-
(*iii*) mentioned above. In particular, the
third feature allows the control to enhance
the parameter identification in real time.
This is done using the control's dual
effect [F1]: the control can affect the

*The strict meaning of Certainty Equivalence
is that all the random variables in the
problem under consideration can be replaced
by their means-the problem is equivalent to
one with perfect certainty.

state as well as the (augmented) state uncertainty and thus can reduce the uncertainty about some parameters. This is the "probing" or "estimation/identification enhancement" property of the control. For this reason the algorithm was also called "dual control." At the same time the control also has to exercise "caution" in order to avoid the performance to suffer due to the existing uncertainties.

The connection between the stochastic Dynamic Programming and these two properties of "probing" and "caution" of an adaptive controller is discussed in Section 2.

A flight control application in which stochastic adaptive control appears to offer significant payoff is the active control of aircraft wing-store flutter. Fighter aircraft are required to carry many different combinations of external wing-mounted stores to perform a variety of missions over a wide operational envelope. Wing mounting of these stores gives rise to different flutter speeds. Release of the wing-mounted stores will cause an abrupt change in the damping and frequencies of wing structural modes. The structural and aerodynamic models used in the design of "constant gain" type controllers are increasingly inaccurate for higher frequency aero-elastic dynamics. Thus, improved flutter suppression could be accomplished with an adaptive controller, which includes the capability to learn and identify the flutter modes during the flight mission.

The ability to successfully suppress flutter during a change in store configuration requires that the adaptive controller identify the structural modes very rapidly. Failure to identify the system parameters quickly enough could result in an instability or cause structural damage. For this reason, an adaptive control which provides identification enhancement through probing would result in more rapid identification of system parameters than a heuristic certainty equivalence controller.

Section 3 describes the flutter model considered and simulation results are presented in Section 4. It is shown that the CL control, by anticipating the learning of the parameter can enhance their identification; i.e., be "actively adaptive." The HCE control is adaptive, but only passively so, and its "accidental learning" is not as fast as the CL controller's.

2. PROBING AND CAUTION IN ADAPTIVE CONTROL

The actively adaptive control approach developed earlier in [T1,T2,B4] is described in this section and a decomposition of the stochastic cost is presented that will indicate the effect of the uncertainties on the control -- whether it should be more aggressive or more cautious in comparison with the heuristic certainty equivalence (HCE - when

all the random variables are replaced by their means). This algorithm is suboptimal in the sense that certain approximations are used in expressing the optimal return function in the solution of the dynamic programming equation. In particular, Taylor's series expansions about some nominal trajectory, including second order terms, are used. The convenient and intuitively appealing form of the solution, together with its computational tractability, however, make it a very useful tool. Only a brief outline of the algorithm is given to facilitate understanding of the stochastic cost decomposition (see [B4] for details).

Consider the system whose state $\underline{x}(k)$, an n-vector, (which has been augmented to include unknown parameters) evolves according to the equation

$$\underline{x}(k+1) = \underline{f}[k,\underline{x}(k), \underline{u}(k)] + \underline{v}(k) \quad (2.1)$$
$$k = 0,1,\ldots N-1$$

and whose observations are given by $\underline{y}(k)$, an m-vector, according to

$$\underline{y}(k) = \underline{h}[k,\underline{x}(k)] + \underline{w}(k), \quad k = 1,\ldots,N-1 \quad (2.2)$$

The initial condition, $\underline{x}(0)$, is a random variable with mean $\hat{\underline{x}}(0/0)$ and covariance $\underline{\Sigma}(0/0)$; $\underline{v}(k)$ and $\underline{w}(k)$ are the process and measurement noises, with known statistics up to second order. The cost function is taken as

$$C(N) = \psi[\underline{x}(N)] + \sum_{k=0}^{N-1} L[\underline{x}(k),k] + \phi[\underline{u}(k),k] \quad (2.3)$$

The optimal closed-loop expected cost-to-go can be written as [B4]

$$J^{CL}(N-k) \overset{\Delta}{=} J_D(N-k) + J_C(N-k) + J_P(N-k) \quad (2.4)$$

where

$$J_D(N-k) \overset{\Delta}{=} \phi[\underline{u}(k),k] + C_0(N-k-1) + \gamma_0(k+1) \quad (2.5)$$

is the deterministic part of the cost and

$$J_C(N-k) \overset{\Delta}{=} \tfrac{1}{2}tr[\underline{K}_0(k+1) \; \underline{\Sigma}(k+1|k)] +$$
$$+ \tfrac{1}{2} \sum_{j=k+1}^{N-1} tr[\underline{K}_0(j+1)\underline{V}(j)] \quad (2.6)$$

$$J_P(N-k) \overset{\Delta}{=} \tfrac{1}{2} \sum_{j=k+1}^{N-1} tr[A_{0,\underline{xx}}(j) \; \Sigma_0(j|j)] \quad (2.7)$$

are the stochastic terms in the cost obtained via the perturbation problem. In the above, V is the process noise covariance, Σ is the covariance of the augmented state and γ, K

and A are given by appropriate recursions detailed in [B4].

The first stochastic term, (2.6), reflects the effect of the uncertainty at time k and subsequent process noises on the cost. These uncertainties cannot be affected by $\underline{u}(k)$ but their weightings do depend on it. The effect of these uncontrollable uncertainties on the cost should be minimized by the control; this term indicates the need for the control to be cautious and thus is called *caution term*. The *second stochastic* term, (2.7), accounts for the effect of uncertainties when subsequent decisions (corrective actions) will be made. The weighting of these future uncertainties is non-negative ($A_{0,xx}$ is positive semidefinite). If the control can reduce by probing (experimentation) the future updated covariances, it can thus reduce the cost. The weighting matrix $A_{0,xx}$ yields approximately the value of future information for the problem under consideration. Therefore this is called the **probing** term. Note that even if the control has no dual effect, i.e., it does not affect the future covariance Σ of the augmented state (which includes the random parameters), the weighting of these covariances is still affected by the control. Therefore this procedure accounts not only for the dual effect but all the stochastic effects in the performance index.

The benefit of probing is weighted by its cost and a compromise is chosen such as to minimize the sum of the deterministic, caution and probing terms. The minimization of J^{CL} will also achieve a tradeoff between the present and future actions according to the information available at the time the corresponding decisions are made.

To find the closed-loop control $\underline{u}(k)$, the minimization of (2.4) is performed using a search procedure. At every k to each control $\underline{u}(k)$ for which (2.4) is evaluated during the search there corresponds a predicted state and to this predicted state a sequence of deterministic controls is attached that defines the nominal trajectory. The only use of the nominals and perturbations is to make possible the evaluation of the cost-to-go optimized in a closed-loop manner. This procedure is repeated at every time a new control is to be obtained.

If the uncertainty dominates the problem then one can distinguish two cases: (1) The caution component, J_C, dominates. Then, since this is "uncontrollable" uncertainty, one has a highly uncertain model which cannot be improved in the course of the control period. (2) The probing component, J_P, dominates. Then, with the dual effect of the control one can reduce the uncertainty of the model - thus the model, while uncertain at the beginning, might prove to be ultimately adequate for the control problem under consideration. A third case occurs when (3). The deterministic component of the cost, I_D, dominates: then the

parameter uncertainties are of no significant consequence. This is the most desirable situation because then we can use CE, i.e., least expensive, control algorithm with good performance. However, **only** the stochastic control approach can indicate this.

3. A SIMPLIFIED WING-STORE FLUTTER MODEL

A simplified version of a wing store flutter model can be represented by a second order differential equation. The state space model, with position and velocity components, can be written as

$$\underline{\dot{x}} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\zeta\omega_0 \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ \kappa \end{bmatrix} u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v \quad (3.1)$$

with measurements of velocity only

$$y = [0 \quad 1] \ \underline{x} + w \quad (3.2)$$

where v and w are the process and measurement noises, respectively.

Typical values of the parameters for model (3.1) are $\omega_0 = 20 \pm 10$, $\zeta = 0.05 \pm 0.1$ (it can become open-loop unstable) and $\kappa = 1 \pm 0.9$ (the control gain can become very low).

A more general flutter model would include a lead-lag transfer function between control input and input u of model (3.1). However, the simplified model (3.1) is sufficient to demonstrate the adaptive control concept of improved control by identification enhancement.

The discretized version of (3.1) is, for sufficiently high sampling rate (typically ten times its natural frequency)

$$\underline{x}(k+1) = \begin{bmatrix} 1 & \Delta T \\ -\omega_0^2 \Delta T & 1-2\zeta\omega_0\Delta T \end{bmatrix} \underline{x}(k)$$

$$+ \begin{bmatrix} 0 \\ \kappa\Delta T \end{bmatrix} u(k) + \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix} \quad (3.3)$$

where v(k) is a zero-mean white noise sequence.

For $\omega_0 = 20$ one has $f = 20/2\pi \approx 3.2$, $T \approx 0.3$ and the sampling time was chosen as $\Delta T = 0.03$. The nominal parameters of the discrete time model are then

$$O_1 = -\omega_0^2 \Delta T = -12$$

$$O_2 = 1-2\zeta\omega_0\Delta T = 0.94 \quad (3.4)$$

$$O_3 = \kappa\Delta T = 0.03$$

The augmented state model consists of (3.3) and the model for the parameters with additive zero-mean white noise

$$O_i(k+1) = O_i(k) + v_{i+2}(k) \quad i=1,2,3 \quad (3.5)$$

i.e. the parameters were assumed to behave (over the relatively short horizon of the problem) as Wiener processes. This was done to allow for the changes that occur in the flutter dynamics during the flight.

The initial estimate for the augmented state was

$$\hat{\underline{x}}(0|0) = [0 \quad 10 \quad -12 \quad .94 \quad 0.03]' \quad (3.6)$$

with the covariance matrix assumed diagonal

$$\Sigma(0|0) = \text{diag}[10^{-2} \quad 1 \quad 36 \quad \sigma_{\Theta_2}^2 \quad \sigma_{\Theta_3}^2] \quad (3.7)$$

The last two terms, reflecting the damping and input gain uncertaintiy, can take a number of values.

The process noise covariance was

$$V = \text{diag} [0, \quad 10^{-2}, \quad 0, \quad V_{44} \quad V_{55}] \quad (3.8)$$

The terms V_{44} and V_{55} were non-zero in the runs where the effects of time-varying damping and control gain, respectively, were investigated.

The flutter control problem can be represented as the minimization of a quadratic cost criterion

$$J = E\left\{ \sum_{k=1}^{N} \underline{x}'(k)Q(k)\underline{x}(k) + ru^2(k-1) \right\} \quad (3.9)$$

with

$$Q(k) = \begin{bmatrix} .001 & 0 \\ 0 & .1 \end{bmatrix} ; \quad r=1 \quad (3.10)$$

where, N is chosen to reflect the desired sample duration during the store configuration change. For the problem here N=5 was chosen. As indicated by (3.10) the goal is to keep the velocity, x_2, small with limited amounts of control.

4. SIMULATION REUSLTS

The flutter model of (3.3) and (3.4) was investigated with nominal parameter values shown in (3.6), (3.7) and (3.8). Two controllers were evaluated: (1) the closed loop control u^{CL} which minimizes the quadratic cost (3.9) and assumes uncertainty in the flutter parameters and (2) the Heuristic Certainty Equivalence control u^{HCE} which assumes the flutter parameters are known without error. The case of time-invariant parameters is shown first, followed by assuming the parameters vary with time (Wiener Process) as shown in (3.5).

The first set of simulations consisted of the evaluation of the first period cost decomposition presented in the previous section for time-invariant parameters. Table 4.1 presents the results in terms of the cost components evaluated at the Heuristic Certainty Equivalence

control value u^{HCE} (0) and at the value obtained by minimizing (2.4), u^{CL} (0). In cases 1 and 2, with moderate uncertainties in Θ_2 (damping) and Θ_3 (input gain) the three cost components - deterministic, probing and caution - are of approximately the same magnitude. The minimum of the closed-loop cost is very close to the HCE control, which minimizes only the deterministic cost (because HCE ignores all uncertainties). For larger uncertainties in the damping the caution component increases but the reduction in the probing component, with a larger magnitude control $|u^{CL}|>|u^{HCE}|$, yields a small reduction of the total cost. Case 5 considers the situation where the gain uncertainty is very large. This situation leads to a significant dominance of the caution effect - the magnitude of the CL control is significantly smaller than the HCE control.

The significance of the results presented in Table 4.1 for the flutter problem (assuming a time-invariant parameter description) is as follows. For large uncertainty in the damping parameter (Θ_2) the performance of both the u^{CL} and the u^{HCE} controller is nearly the same. However, if in addition, the control gain has large uncertainty (case 5) the CL control shows a 7% reduction in the cost - all of which is due to the caution component. This implies that uncertain knowledge of the control gain dictates that the optimal control should exhibit more caution than the Heuristic Certainty Equivalence controller.

The second set of simulations was performed for a time-varying description of the flutter parameters. A time-varying parameter case was simulated by assuming there is process noise in (3.5) for i=3. The standard deviation of the noise affecting the input gain was taken as $\sqrt{V_{55}} = 0.014$. The results are shown in Table 4.2 for different values of the initial damping uncertainty. As can be seen probing dominates: a significant reduction in the probing cost and a 10% reduction in the total cost can be obtained by using an actively adaptive control like u^{CL}. This control anticipates that changes will occur in the parameter even though it does not know what will be the changes, which are modelled by zero-mean noise with variance V_{55}, according to (3.5). Consequently, this "anticipation" (which is restricted to be causal) leads the control to enhance the identification of the input gain, whose variance otherwise would be excessively large.

The results in Table 4.2 demonstrate that flutter suppression can be more effectively achieved by probing the system to enhance identification of the control gain for the case where the control gain can vary with time. The results of Table 4.2 indicate the average performance improvements by using the CL-controller. Specific time history results can give a detailed examination of the identification enhancement property of the CL-control.

The next set of simulations consists of time history runs with time-varying parameter as in case 7. The true value for the gain was $O_3=0.03$. The process noise $v_5(1)$ simulated the change of O_3 from time 1 to time 2. The goal was to see how the probing control as shown in case 7 (Table 4.2) was able to enhance the real-time parameter identification in order to reduce the cost. An exact assessment of the potential benefits from using u^{CL} vs. u^{HCE} would involve many Monte Carlo runs where all the random variables (initial conditions, parameters, noises) have to be generated according to their statistical characterizations [B5] and the results require special analysis [W1]. A few runs only were carried out with only the noise $v_5(1)=O_3(2)-O_3(1)$ being non-zero while, all the other noises were set to zero, to evaluate the cumulated cost over N=5 steps. Table 4.3 shows these values for the two control policies for a few parameter changes. In cases 8-10 the initial estimate of the input gain was the same as the true value, i.e., $\hat{O}_3(0) = O_3(0) = 0.03$. In this situation, which initially favors the HCE controller, the CL controller is still better when the gain decreases (cases 9 and 10). Note that this decrease of the control gain causes significant cost increases and this is when the CL controller proves itself useful. In cases 11 and 12 the initial gain estimate was $\hat{O}_3(0) = 0.05$, i.e., it was overestimated.

The final set of simulations represent time histories where both the damping parameter O_2 and the control gain O_3 experience abrupt changes. This would be typical of a wing store configuration change. The damping and control gain change are shown in Fig. 4.1.

For this case the damping parameter (O_2) goes from a stable value of .94 to an unstable value of 1.06. The control gain (O_3) goes from .03 to .005. The standard deviation of the noise for the damping parameter was $\sqrt{V_{44}} = .1$.

The cumulated cost for this case is shown in Table 4.4 for u^{CL} and u^{HCE}.

The CL control is seen to have improved the performance over the HCE controller. This improved performance is due to identification enhancement of the damping parameter. This can be seen in Figure 4.2 where the CL control is shown to identify the damping parameter more rapidly. Figure 4.3 shows the identified gain parameter which is successfully identified by both controllers after 5 time steps.

5. CONCLUSION

The simulation results presented in this paper indicate that potential improvement in flutter suppression is possible using an adaptive control of the closed loop type. This improvement is a direct result of identification enhancement due to probing in

the control solution. A more detailed flutter model and further simulation is required to fully quantify the maximum achievable performance capability using the CL control.

REFERENCES

[B1] R. Bellman, Dynamic Programming, Princeton University Press, 1957.

[B2] R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton NJ: Princeton University Press, 1961.

[B3] Y. Bar-Shalom and E. Tse, "Dual Effect, Certainty Equivalence, and Separation in Stochastic Control", IEEE Transactions on Automatic Control, AC-19, pp. 494-500, October 1974.

[B4] Y. Bar-Shalom and E. Tse, "Concepts and Methods in Stochastic Control", in C. T. Leondes (Ed.) Control and Dynamic Systems: Advances in Theory and Application, Vol. 12, Academic Press, 1976.

[B5] Y. Bar-Shalom, "Stochastic Dynamic Programming, Caution and Probing," Proc. Joint Automatic Control Conf., San Francisco, CA, Aug. 1980.

[F1] A. A. Feldbaum, Optimal Control Systems, New York: Academic Press, 1965.

[S1] G. Saridis, Self-Organizing Control of Stochastic Systems, Marcel Dekker, 1976.

[T1] E. Tse, Y. Bar-Shalom and L. Meier, "Wide-sense Adaptive Dual Control of Stochastic Nonlinear Systems," IEEE Trans. Automatic Control, Vol. AC-18, pp. 98-108, April 1973.

[T2] E. Tse and Y. Bar-Shalom, "An Actively Adaptive Control for Discrete-Time Systems with Random Parameters", IEEE Transactions on Automatic Control, Vol. AC-18, pp. 109-117, April 1973.

[W1] C. J. Wenk and Y. Bar-Shalom, "A Multiple Model Adaptive Dual Control Algorithm for Stochastic Systems with Unknown Parameters," IEEE Trans. Automatic Control, Vol. AC-25, Aug. 1980.
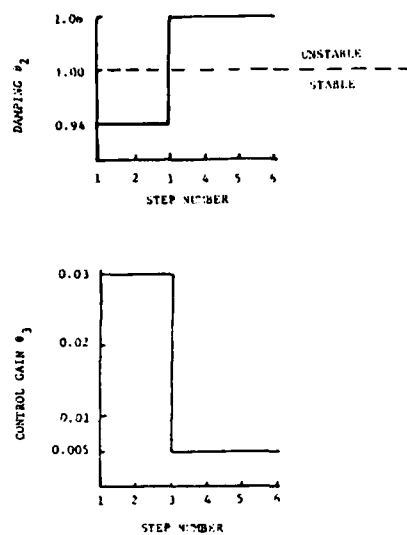
Figure 4.1. Time history simulation of damping and control gain to represent wing store flutter configuration change.

Case	z_{θ_2}	σ_{θ_3}	Control	J^{CL}	J_D	J_P	J_C
1	0.02	0.01	u^{HCE}=-.87	2.951	0.735	1.371	0.845
2	0.1	0.01	u^{HCE}=-.287	3.006	0.735	1.375	0.896
3a	0.5	0.01	u^{HCE}=-287	4.33	0.735	1.406	2.188
3b	0.5	0.01	u^{CL}=-304	4.31	0.751	1.359	2.202
4a	1.0	0.01	u^{HCE}=-287	8.377	0.735	1.41	4.225
4b	1.0	0.01	u^{CL}=-332	8.236	0.846	1.24	6.094
5a	0.5	0.03	u^{HCE}=-287	9.24	0.735	1.526	6.98
5b	0.5	0.03	u^{CL}=-215	8.648	1.024	1.879	5.739

Table 4.1 First period cost decompositions for time-invariant parameters

Case	σ_{θ_2}	σ_{θ_3}	Control	J^{CL}	J_D	J_P	J_C
6a	0.01	0.01	u^{HCE}=-287	6.41	0.735	4.27	1.406
6b	0.01	0.01	u^{CL}=-360	5.85	1.029	3.37	1.45
7a	0.5	0.01	u^{HCE}=-287	7.738	0.735	4.303	2.698
7b	0.5	0.01	u^{CL}=-380	7.06	1.213	3.18	2.66

Table 4.2 First period cost decomposition for time-varying input gain
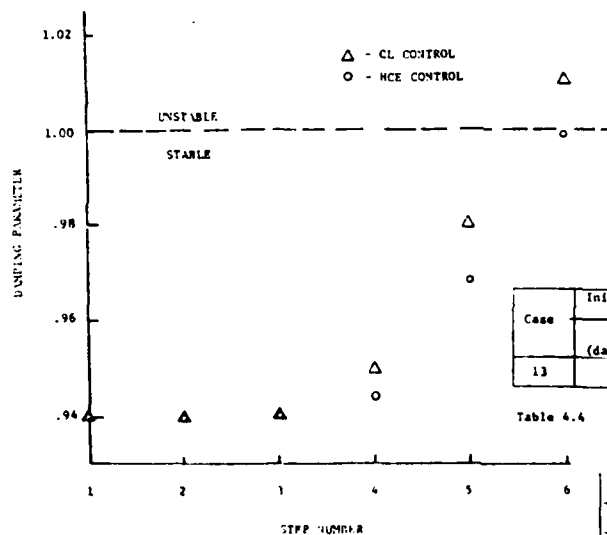
Figure 4.2 Identified damping parameter for Case 13.

Case	Initial Parameter Estimate		Parameter Change		Cumulated Cost	
	θ_2 (damping)	θ_3 (gain)	θ_2 (damping)	θ_3 (gain)	HCE	CL
13	.94	.03	1.06	.005	130.1	114.6

Table 4.4 Time history runs with both input gain change and damping parameter change

Case	Initial input gain estimate	Input gain change	Cumulated cost	
			CE	CL
8	0.03	0.02	0.96	1.3
9	0.03	-0.02	3.59	3.34
10	0.03	-0.025	10.99	7.49
11	0.05	-0.02	7.38	6.56
12	0.05	-0.025	20.38	17.7

Table 4.3 Time history runs with input gain change
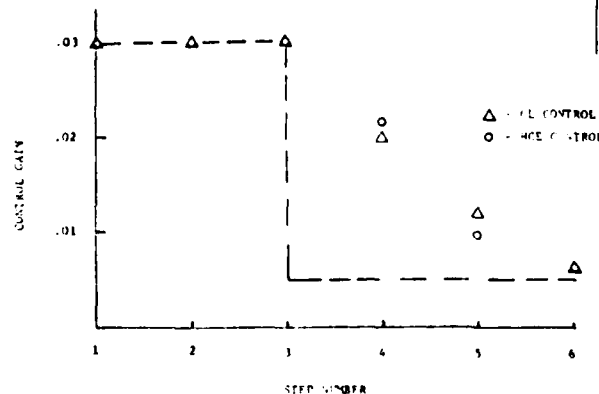
Figure 4.3 Identified control gain for Case 13

A LINEAR FEEDBACK DUAL CONTROLLER
FOR A CLASS OF STOCHASTIC SYSTEMS

Y. Bar-Shalom, P. Mookerjee and J. A. Molusis

University of Connecticut
Department of Electrical Engineering
and Computer Science
Storrs, CT, USA 06268

Résumé.

On présente une méthode pour la construction d'un algorithme de commande
duale ayant une structure à rétroaction linéaire. L'application de cet
algorithme pour la commande d'un hélicoptère est discutée et des résultats
de simulation sont donnés.

Abstract

The methodology for deriving a dual control algorithm that has a linear
feedback form is presented. This control, while simple, has the capability
of enhancing the identification of the system's unknown parameters. A dual
controller for a plant describing the helicopter higher harmonic vibration
control problem is presented together with simulation results.

1.   Introduction

In the control of nonlinear stochastic systems the control has, in
general, a dual effect [F1, B1]: it affects the system's state as well as
its uncertainty. Since in linear plants with unknown parameters the con-
trol has a dual effect, it can be potentially used to enhance the real-time
identification of the system parameters.

The attractiveness of a linear controller that incorporates the dual
effect has been pointed out in [M1]. Previous dual control algorithms
[A2, B2, W1, W2] required numerical search which makes their implementation
costly. The success of the self-tuning regulator [A1], which stems from
its ease of implementation as well as its effectiveness, prompted us to
investigate control algorithms that have a linear feedback form but incor-
porate the dual effect.

The problem considered in Section 2 is the simplest one where there
is a dual effect, in order to illustrate the concept. A 2-stage optimi-
zation problem is then formulated with the stochastic dynamic programming
in Section 3 and the controller is derived in Section 4.

In Section 5 an algorithm based on this methodology is derived for
a multiple-input multiple-output model corresponding to a simplified
version of the "higher harmonic control" of helicopter vibration [W3, M2].
Simulation results are presented in Section 6.

2. Problem Formulation

The following memoryless unknown-gain system with plant and measurement noises is considered. The plant equation is

$$x(k+1) = bu(k) + v(k) \tag{2.1}$$

with

$$Ev(k) = 0 \quad ; \quad Ev(k)v(j) = V \delta_{kj} \tag{2.2}$$

and the measurement is given by

$$y(k) = x(k) + w(k) \tag{2.3}$$

where

$$Ew(k) = 0 \quad ; \quad Ew(k)w(j) = W \delta_{kj} \tag{2.4}$$

and

$$Ev(k)w(j) = 0 \tag{2.5}$$

The estimation of the unknown gain b (assumed time invariant here) is done according to the following equations:

$$\hat{b}(k+1) = \hat{b}(k) + P(k)u(k) [P(k)u^2(k) + V + W]^{-1}[y(k+1) - \hat{b}(k)u(k)] \tag{2.6}$$

$$P(k+1) = E[b - \hat{b}(k+1)]^2 = P(k)(V + W) [P(k)u^2(k) + V + W]^{-1} \tag{2.7}$$

Note in (2.7) the fact the control affects the variance of the parameter estimate, i.e., it has the dual effect [F1, B1].

The control criterion to be minimized will be taken as the expected value of the cost from step 0 to N

$$J(0) = E\{C(0)\} \tag{2.8}$$

where

$$C(k) = \sum_{j=k}^{N} c[j,x(j),u(j)] \tag{2.9}$$

and, with $\xi(j)$ denoting the desired state at time j

$$c(j) = q(j)[x(j) - \xi(j)]^2 + ru^2(j) \quad j=0,1,\ldots,N-1 \tag{2.10}$$

$$c(N) = q(N)[x(N) - \xi(N)]^2 \tag{2.11}$$

3. The Multistage Problem and Dynamic Programming

The general equation of the Stochastic Dynamic Programming is

$$J^*(k,Y^k) = \min_{u(k)} E[c(k) + J^*(k+1, Y^{k+1})|Y^k] \quad k=N-1,\ldots,1,0 \tag{3.1}$$

where $J^*(k)$ is the "cost-to-go" from k to N, Y^k is the cumulated information at time k when the control u(k) is to be determined.

Due to the memoryless nature of the system (2.1) the only coupling between the stages in a multistage problem is the informational effect of the control - its effect on the quality of the estimate of the parameter b.

The last control is obtained from

$$J^*(N-1) = \min_{u(N-1)} E\{q(N-1)[x(N-1)-\xi(N-1)]^2+ru^2(N-1)+q(N)[x(N)-\xi(N)]^2 | Y^{N-1}\}$$

$$= \min_{u(N-1)} \{E[q(N-1)[x(N-1)-\xi(N-1)]^2|Y^{N-1}]+[r+q(N)[\hat{b}^2(N-1) +$$

$$+ P(N-1)]]u^2(N-1)-2q(N)\xi(N)\hat{b}(N-1)u(N-1)+q(N)[V+\xi^2(N)]\} \qquad (3.2)$$

as

$$u^*(N-1) = \{r+q(N)[\hat{b}^2(N-1)+P(N-1)]\}^{-1}q(N)\xi(N)\hat{b}(N-1) \qquad (3.3)$$

This yields the optimal cost-to-go

$$J^*(N-1) = E[q(N-1)(x(N-1) - \xi(N-1))^2|Y^{N-1}] + \tilde{J}^*(N-1) \qquad (3.4)$$

where

$$\tilde{J}^*(N-1) = -[r+q(N)(\hat{b}^2(N-1)+P(N-1))]^{-1}(q(N)\xi(N)\hat{b}(N-1))^2+q(N)\xi^2(N)+q(N)V$$

$$= [r+q(N)(\hat{b}^2(N-1)+P(N-1))]^{-1}(r+q(N)P(N-1))q(N)\xi^2(N)+q(N)V \qquad (3.5)$$

is the cost-to-go excluding the term which is not affected by the current control.

The control (3.3) is the well-known "one step ahead cautious" control. This is the optimal control, for all k, if the cost has a sliding horizon of only one step (called also "myopic" control).

The next to the last control is to be obtained from the following

$$J^*(N-2,Y^{N-2}) = \min_{u(N-2)} E\{c(N-2)+J^*(N-1,Y^{N-1})|Y^{N-2}\} \qquad (3.6)$$

The dependence of $J^*(N-1)$, given by (3.4), on $y(N-1)$ is via $\hat{b}(N-1)$. Since, as detailed in (3.5), $J^*(N-1)$ is a rational function of $\hat{b}(N-1)$ one cannot carry out explicitly the expectation in (3.6), which is over $y(N-1)$ conditioned on Y^{N-2}. Even if one could carry out explicitly this expectation, the dependence of the cost-to-go $J^*(N-1)$ on the previous control $u(N-2)$ via $P(N-1)$ poses a significant problem: the minimization of (3.6) would require solving a high order algebraic equation. This can be seen as follows.

Assume that $\hat{b}(N-1)$ in $\hat{J}(N-1)$ given by (3.4), (3.5) would be replaced by $\hat{b}(N-2)$, the estimate at the time $u(N-2)$ is to be computed. This removes the need to carry out the expectation of $J^*(N-1)$ conditioned on Y^{N-2} in (3.6). Then (3.6) becomes an explicit function of $u(N-2)$ and, as shown in Sternby [S1], the derivative w.r.t. $u(N-2)$ leads to a fifth order polynomial.

Thus the two main problems in performing the first backward iteration of the Stochastic Dynamic Programming as given in (3.6) are the conditional expectation over the future measurement and the minimization. In the Linear-Quadratic Problem the presence of quadratic and linear terms (as opposed to rational functions here) made possible an easy solution for the optimal control. The resulting solution, in the form of a linear feedback control has been in wide usage because of its ease of implementation. On

the other hand, the linear problem with unknown parameters is encountered in many applications and it is desirable to obtain (and evaluate) a dual controller which has the linear feedback form. The gain should in this case depend on the current as well as the expected future parameter uncertainties.

4.   A Linear Feedback Dual Controller with a Two-Step Horizon

The cost-to-go given in (3.5) depends on the following variables:

$$\tilde{J}^*(N-1) = \tilde{J}^*[N-1, \hat{b}^2(N-1), P(N-1)] \qquad (4.1)$$

The first, $\hat{b}^2(N-1)$, the estimate squared of the parameter at N-1, will have to be "averaged out" conditioned on Y^{N-2}. The second, $P(N-1)$ depends directly on $u(N-2)$, which is to be determined from (3.6).

The following first order series expansion of (4.1) is proposed

$$\tilde{J}^*(N-1) \simeq \tilde{J}^*[N-1, \hat{b}^2(N-2), \bar{P}(N-1) + \frac{\partial \tilde{J}^*(N-1)}{\partial \hat{b}^2(N-1)} [\hat{b}^2(N-1) - \hat{b}^2(N-2)]$$

$$+ \frac{\partial \tilde{J}^*(N-1)}{\partial P(N-1)} \frac{\partial P(N-1)}{\partial u^2(N-2)} [u^2(N-2) - \bar{u}^2(N-2)] \qquad (4.2)$$

In other words, the expansion is about the current estimate of the parameter, $\hat{b}(N-2)$, and a "nominal" variance for this parameter $\bar{P}(N-1)$, given by

$$\bar{P}(N-1) = P(N-2)(V+W)[P(N-2)\bar{u}^2(N-2) + V + W]^{-1} \qquad (4.3)$$

where $\bar{u}(N-2)$ is a "nominal" control at N-2.

The following notations are introduced

$$\bar{J}(N-1) \triangleq \tilde{J}^*[N-1, \hat{b}^2(N-2), \bar{P}(N-1)] \qquad (4.4)$$

$$\bar{J}_b(N-1) \triangleq \frac{\partial \tilde{J}^*(N-1)}{\partial \hat{b}^2(N-1)} \qquad (4.5)$$

$$\bar{J}_P(N-1) \triangleq \frac{\partial \tilde{J}^*(N-1)}{\partial P(N-1)} \qquad (4.6)$$

$$\bar{P}_u(N-1) \triangleq \frac{\partial P(N-1)}{\partial u^2(N-2)} \qquad (4.7)$$

where the partial derivatives (4.5)-(4.7) are evaluated at the same nominal values as (4.4). Note that (4.5) and (4.6) are the sensitivities of the cost-to-go w.r.t. the parameter and its uncertainty, respectively; (4.7) is the sensitivity of the parameter uncertainty w.r.t. the control. With these notations (4.2) can be written

$$\tilde{J}^*(N-1) = \bar{J}(N-1) + \bar{J}_b(N-1) \cdot [\hat{b}^2(N-1) - \hat{b}^2(N-2)] + \bar{J}_p(N-1)\bar{P}_u(N-1) \cdot$$

$$\cdot [u^2(N-2) - \bar{u}^2(N-2)] \tag{4.8}$$

The asterisk on the cost, symbolizing optimality, has been kept even though (4.8) is only an approximation to the optimum.

When inserting (4.8) into (3.6) its expected value conditioned on Y^{N-2} will have to be computed. Note that only the second term on the r.h.s. of (4.8) is random when conditioned on Y^{N-2}. Its conditional expectation is

$$E\{\bar{J}_b(N-1) \ [\hat{b}^2(N-1) - \hat{b}^2(N-2)]|Y^{N-2}\}$$

$$= \bar{J}_b(N-1) \ [\ E \ \{\hat{b}^2(N-1)|Y^{N-2}\} - \hat{b}^2(N-2)]$$

$$= \bar{J}_b(N-1) \ [\ P(N-2) - P(N-1)] \tag{4.9}$$

Notice the fact that u(N-2) enters into (4.9) via P(N-1). A first order expansion of P(N-1) about its nominal value (4.3) will be used in (4.9). Using notation (4.7) one replaces (4.9) by

$$E \ \{\bar{J}_b(N-1) \quad \cdot \quad [\hat{b}^2(N-1) - \hat{b}^2(N-2)] \ | \ Y^{N-2}\}$$

$$\approx \bar{J}_b(N-1) \cdot [P(N-2) - \bar{P}(N-1) - \bar{P}_u(N-1) \cdot [u^2(N-2) - \bar{u}^2(N-2)] \tag{4.10}$$

The (approximate) conditional mean of (4.8), becomes, using (4.10)

$$E \ \{\tilde{J}^*(N-1) \ | \ Y^{N-2}\} = \bar{J}(N-1) + \bar{J}_b(N-1) \cdot [P(N-2) - \bar{P}(N-1)]$$

$$+ [\bar{J}_p(N-1) - \bar{J}_b(N-1)] \ \bar{P}_u(N-1) \cdot [u^2(N-2) - \bar{u}^2(N-2)] \tag{4.11}$$

Combining (4.11) and (3.4) into (3.6) yields

$$J^*(N-2) = \min_{u(N-2)} \ \{E[q(N-2)(x(N-1) - \xi(N-1))^2 + ru^2(N-2) + q(N-1)(x(N-1) -$$

$$- \xi(N-1))^2| \ Y^{N-2}] + \bar{J}(N-1) + \bar{J}_b(N-1)[P(N-2) - \bar{P}(N-1)] +$$

$$+ [\bar{J}_p(N-1) - \bar{J}_b(N-1)] \ \bar{P}_u(N-1)[u^2(N-2) - \bar{u}^2(N-2)]\} \tag{4.12}$$

Ignoring the terms in (4.12) that are independent of u(N-2) yields

$$u^*(N-2) = \text{arg min} \ \{q(N-1)E[(x(N-1) - \xi(N-1))^2 \ | \ Y^{N-2}]$$

$$+ [r + (\bar{J}_p(N-1) - \bar{J}_b(N-1))\bar{P}_u(N-1)] \ u^2(N-2)\}$$

$$= \text{arg min} \ \{[r+q(N-1)[\hat{b}^2(N-2)+P(N-2)]+[\bar{J}_p(N-1)-\bar{J}_b(N-1)]\bar{P}_u(N-1)]u^2(N-2)$$

$$- 2q(N-1)\xi(N-1)\hat{b}(N-2)u(N-2)\} \tag{4.13}$$

which gives the control as

$$u^*(N-2) = [r + q(N-1)(\hat{b}^2(N-2) + P(N-2) + (\bar{J}_p(N-1) - \bar{J}_b(N-1))\bar{P}_u(N-1)]^{-1}$$

$$q(N-1)\xi(N-1)\hat{b}(N-2) \tag{4.14}$$

Note the presence of the caution effect above - the additive $P(N-2)$ in the denominator, which being positive, will tend to <u>decrease</u> the control magnitude. However, the last term in the denominator is negative reflecting the probing effect via the sensitivity functions (4.5 - 4.7) and thus will tend to <u>increase</u> the control. This can be seen as follows: $\bar{J}_p(N-1)$ is positive (since the cost increases with uncertainty), $\bar{J}_b(N-1)$ is negative (this follows from inspection of (3.5)), and $\bar{P}_u(N-1)$ is negative (this follows from inspection of (2.7)).

The resulting control has thus the linear feedback form with the gain modified by the caution <u>and</u> probing effects.

5. Extension to Multiple Input Multiple Output Model

The plant model is

$$\underline{x}(k+1) = \underline{c} + B\,\underline{u}(k) + \underline{v}(k) \tag{5.1}$$

with

$$E\,\underline{v}(k) = 0 \quad ; \quad E\,\underline{v}(k)\underline{v}'(j) = V\,\delta_{kj} \tag{5.2}$$

where $\underline{c}$ is an unknown vector, B a matrix with unknown parameters. The unknown elements of $\underline{c}$ and B are denoted as $\underline{\theta}$ with covariance matrix P. In the helicopter vibration problem to be considered later $\underline{c}$ is the amplitude of uncontrolled vibrations. The matrix B is called the "transfer matrix" [M2] and represents the effect of the control on the vibration amplitude.

The measurement is given by

$$\underline{y}(k) = \underline{x}(k) + \underline{w}(k) \tag{5.3}$$

where

$$E\,\underline{w}(k) = 0 \quad ; \quad E\,\underline{w}(k)\underline{w}'(j) = W\,\delta_{kj} \tag{5.4}$$

$$E\,\underline{v}(k)\underline{w}'(j) = 0$$

The control criterion to be minimized is the expected value of the cost from step 0 to N

$$J(0) = E\{C(0)\} = E\{\sum_{k=1}^{N} \underline{x}'(k)Q\underline{x}(k) + \underline{u}'(k-1)R\underline{u}(k-1)\} \tag{5.6}$$

The last control is easily obtained by minimizing $J(N-1)$ and is given

$$\underline{u}^*(N-1) = -[R + E(B'QB|Y^{N-1})]^{-1}\,E(B'Q\underline{c}|Y^{N-1}) \tag{5.7}$$

Thus inserting $\underline{u}^*(N-1)$ in the cost we get

$$J^*(N-1) = E(\underline{c}'Q\underline{c}|Y^{N-1}) + tr(QV)$$

$$- E(\underline{c}'QB|Y^{N-1})[R + E(B'QB|Y^{N-1})]^{-1} E(B'Q\underline{c}|Y^{N-1}) \tag{5.8}$$

Thus

$$J^*(N-1) = J^*[N-1, \hat{\underline{\theta}}(N-1), P(N-1)] \tag{5.9}$$

where $P(N-1)$ is the covariance matrix associated with the estimate $\hat{\theta}(N-1)$.

The approximation of the stochastic dynamic programming for $N=2$ steps is done with a first order expansion with respect to $\underline{u}(0)$

$$J^*(1) = J^*[1, \hat{\underline{\theta}}(0), \bar{P}(1)] + \frac{\partial J^*(1)}{\partial P(1)} \frac{\partial P(1)}{\partial \underline{u}(0)} \cdot [\underline{u}(0) - \bar{\underline{u}}(0)] \tag{5.10}$$

where

$$J_P(1) \triangleq \frac{\partial J^*(1)}{\partial P(1)} \quad ; \quad P_{\underline{u}}(1) \triangleq \frac{\partial P(1)}{\partial \underline{u}(0)} \tag{5.11}$$

are evaluated at nominal value $\bar{\underline{u}}(0)$.

Then with

$$\bar{J}_1 = J^*[1, \hat{\underline{\theta}}(0), \bar{P}(1)] \tag{5.12}$$

we get

$$J^*(0) = \min_{\underline{u}(0)} E\{\underline{x}'(1)Q\underline{x}(1) + \underline{u}'(0) R\underline{u}(0)|Y^0\}$$

$$+ \bar{J}_1 + J_P(1) P_{\underline{u}}(1) \cdot [\underline{u}(0) - \bar{\underline{u}}(0)]$$

$$= \min_{\underline{u}(0)} E\{[\underline{c} + B\underline{u}(0) + \underline{v}(0)]'Q[\underline{c} + B\underline{u}(0) + \underline{v}(0)] + \underline{u}'(0)R\underline{u}(0) \cdot$$

$$+ \bar{J}_1 + J_P(1) P_{\underline{u}}(1) \cdot [\underline{u}(0) - \bar{\underline{u}}(0)] \mid Y^0\} \tag{5.13}$$

The two-step dual control is then

$$\underline{u}(0) = - [R + E(B'QB|Y^0)]^{-1} [E(B'Q\underline{c}|Y^0) + \frac{\beta}{2} J_P(1) \otimes P_{\underline{u}}(1)] \tag{5.14}$$

where

$$J_P(1) \otimes P_{\underline{u}}(1) = \sum_{m,n} \frac{\partial J_1(1)}{\partial P_{m,n}(1)} \cdot \frac{\partial P_{m,n}(1)}{\partial \underline{u}(0)} \tag{5.15}$$

and $P_{m,n}$ is the m,n element of the matrix P. The coefficient β introduced in (5.14) allows the same expression to yield several controllers as follows:

$\beta = 0$: One step stochastic controller (cautious myopic)

$\beta = 1$: Two step dual controller

$\beta \geq 2$: Modified dual controller with (artificial) extra horizon.

6.   Application to a Helicopter Vibration Control Problem

The problem of helicopter vibration control is to find suitable higher harmonic control amplitudes which, when applied to the system, cancel out

the vibration occurring in the airframe. The relationship between vibration output and higher harmonic control input is known to be nonlinear and thus adaptive control solutions are required. In such cases fixed gain feedback controllers perform poorly. A simplified linear version of this problem (for two vibration components) can be represented by the plant equations [W3]

$$x_1(k+1) = \theta_1 + \theta_2 u_1(k) + \theta_3 u_2(k) + v_1(k)$$

$$x_2(k+1) = \theta_4 + \theta_5 u_1(k) + \theta_6 u_2(k) + v_2(k) \tag{6.1}$$

with

$$E \underline{v}(k) \underline{v}'(k) = V = \text{diag} (V_1, V_2) ; \quad V_1 = 28^2 , V_2 = 440^2 \tag{6.2}$$

The first state, x_1 represents the rotor hub force amplitude at a given frequency (one of the harmonics of the rotor r.p.m.), the second state, x_2, represents the rotor blade bending moment amplitude at the same frequency. The two controls are the "higher harmonic controls". These controls excite the rotor blades at higher harmonics of rotational speed. These cancel out some of the existing unsteady air loads [C1].

The measurements are

$$y_1(k) = x_1(k) + w_1(k)$$

$$y_2(k) = x_2(k) + w_2(k) \tag{6.3}$$

with

$$E \underline{w}(k) \underline{w}'(k) = W = \text{diag} (W_1, W_2) ; \quad W_1 = 28^2 , W_2 = 440^2 \tag{6.4}$$

The initial parameter estimates are generated as $N(\theta_1, \theta_i^2)$, $i = 1, \ldots, 6$ where the true values are

$$\theta_1 = 287.3 \qquad \theta_4 = 4410$$

$$\theta_2 = -25.1 \qquad \theta_5 = -32.5$$

$$\theta_3 = 14.4 \qquad \theta_6 = -54.0 \tag{6.5}$$

The cost weighting matrices are

$$Q = \text{diag} (q_1, q_2) ; \quad q_1 = 10^{-5} , q_2 = 5 \times 10^{-8}$$

$$R = \text{diag} (r_1, r_2) ; \quad r_1 = 10^{-4} , r_2 = 10^{-4} \tag{6.6}$$

In terms of the notation of Section 5

$$\underline{c} = \begin{bmatrix} \theta_1 \\ \theta_4 \end{bmatrix} , \quad B = \begin{bmatrix} \theta_2 & \theta_3 \\ \theta_5 & \theta_6 \end{bmatrix} , \quad \underline{u}(k) = \begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix} \tag{6.7}$$

The parameter vector to be estimated is

$$\underline{\theta}(k) \stackrel{\Delta}{=} [\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_5 \quad \theta_6]' \tag{6.8}$$

and it is modelled as time invariant

$$\underline{\theta}(k+1) = \underline{\theta}(k) \tag{6.9}$$

with measurements

$$y_1(k) = H(k) [\theta_1 \quad \theta_2 \quad \theta_3]' + v_1(k) + w_1(k)$$

$$y_2(k) = H(k) [\theta_4 \quad \theta_5 \quad \theta_6]' + v_2(k) + w_2(k) \tag{6.10}$$

where

$$H(k) = [1 \quad u_1(k) \quad u_2(k)] \tag{6.11}$$

In view of (6.2) and (6.4) the covariance matrix of $\hat{\underline{\theta}}(k)$ is block diagonal

$$P(k) = \begin{bmatrix} P_1(k) & 0 \\ 0 & P_2(k) \end{bmatrix} \tag{6.12}$$

The optimum cost for stage 1, assuming it is the last one, is

$$J^*(1) = E(\underline{c}'Q\underline{c} \mid Y^1) + tr(QV)$$

$$- E(\underline{c}'QB \mid Y^1) [R + E(B'QB|Y^1)]^{-1} E(B'Q\underline{c}|Y^1) \tag{6.13}$$

The above can be rewritten as

$$J^*(1) = q_1(\hat{\theta}_1^2 + P_{1,1}(1)) + q_2(\hat{\theta}_4^2 + P_{4,4}(1)) + q_1 \cdot V_1 + q_2 \cdot V_2$$

$$- \frac{1}{CD-E^2} \ (F^2D - 2FGE + G^2C) \tag{6.14}$$

where

$$C = q_1(\hat{\theta}_2^2 + P_{2,2}(1)) + q_2(\hat{\theta}_5^2 + P_{5,5}(1)) + r_1$$

$$D = q_1(\hat{\theta}_3^2 + P_{3,3}(1)) + q_2(\hat{\theta}_6^2 + P_{6,6}(1)) + r_2$$

$$E = q_1(\hat{\theta}_2\hat{\theta}_3 + P_{2,3}(1)) + q_2(\hat{\theta}_5\hat{\theta}_6 + P_{5,6}(1))$$

$$F = q_1(\hat{\theta}_1\hat{\theta}_2 + P_{1,2}(1)) + q_2(\hat{\theta}_4\hat{\theta}_5 + P_{4,5}(1))$$

$$G = q_1(\hat{\theta}_1\hat{\theta}_3 + P_{1,3}(1)) + q_2(\hat{\theta}_4\hat{\theta}_6 + P_{4,6}(1)) \tag{6.15}$$

The terms $J_p(1)$ are easily obtained from equation (6.14). The covariance update equation is

$$P_i(k) = P_i(k-1) - P_i(k-1)H'(k)[H(k)P_i(k-1)H'(k)+V_i+W_i]^{-1} H(k)P_i(k-1) \tag{6.16}$$

$$i=1,2$$

The nominal covariance $\bar{P}_i(k)$ is obtained in terms of previous $P_1(k-1)$ and a nominal control $\underline{\bar{u}}(k-1)$ of the "1 step" type.

The sensitivity term $P_{\underline{u}}(1)$ can be evaluated from the above.

The two-step dual control (5.14) was implemented for the above

problem with a "sliding horizon" for a total of 20 steps. The evaluation criterion is

$$\sum_{k=1}^{N} \underline{x}'(k)Q\underline{x}(k)$$

Performance was evaluated from 100 Monte Carlo runs for the following cases:

1. Heuristic Certainty Equivalence,
2. One step ahead optimal stochastic cautious myopic,
3. Two step dual
4. Modified two step dual.

The above runs were made for the case $\hat{\theta}_i(0) \sim N(\theta_i, \theta_i^2)$.

Comparisons are made between the performances of the cautious and dual algorithms on the system and a conventional statistical significance analysis is done using the normal theory approach [N1,W1]. The methodology is given in Appendix A. Tables I & II contain the results of the simulation runs. Table II indicates that the dual control performs better than the other controllers over 10 time steps. Table I provides a rigorous argument that the dual outperforms the other controllers.

The performances are compared in Figures 1-3. In Fig. 1 the HCE controller uses a very large control magnitude and drives the system hard. Thus in step 1 the vibration is increased compared to the cautious and dual controllers. This however helps to learn the parameters faster and reduces the vibration earlier than the others. In a realistic situation one cannot really live with a HCE because of the practical bounds on the control. The dual starts off higher than the cautious but behaves better after 2 steps.

Fig. 2 compares the cautious, dual and modified dual algorithms. As β increases from 0 to 6 the vibration at step 1 increases. Values of β from 3 onwards do not behave very much better than $\beta=2$ beyond step 3. Thus $\beta=0,1,2$ are suggested for implementation and the statistical tests were performed only for these values. Fig. 3 compares the cautious and dual over a wider scale.

Single Time History Runs

Results of single time history runs over 20 time steps are plotted in Figs. 4-6 for the HCE, dual and cautious controllers. Fig. 4, 5, and 6 compare the controls U1, U2, and cost for the three cases respectively. For all the controllers the controls U1, U2 reach almost the same value at

the end of 20 steps, although they start differently indicating that the algorithms have learned the parameters. As a trade-off between the rapid learning and smaller cost, the dual is the best of the three.

Algorithms Compared

Cautious myopic – Dual ($\beta=1$)			Cautious myopic – Dual ($\beta=2$)		
Time Step k	Test Statistic z_k	Estimated Improvement $EI_k(\%)$	Time Step k	Test Statistic z_k	Estimated Improvement $EI_k(\%)$
1	-2.30	-7.19	1	-2.83	-20.64
2	-0.36	-1.90	2	-0.21	- 3.2
3	1.26	4.79	3	0.37	3.19
4	5.28	19.56	4	5.32	32.22
5	3.53	23.21	5	7.94	44.48
6	5.43	34.20	6	6.49	47.63
7	4.40	32.51	7	5.43	40.67
8	3.68	34.16	8	4.53	40.67
9	2.94	29.16	9	3.84	36.04
10	2.13	23.39	10	2.81	28.60

Table I. Statistical significance test for algorithm comparisons in the Example (100 Monte Carlo runs)

k	Average Cost over 100 runs							
	$\beta=0$		$\beta=1$		$\beta=2$		HCE	
	$\bar{C}_k^{(1)}$	$\sum_{i=1}^{k} \bar{C}_i^{(1)}$	$\bar{C}_k^{(2)}$	$\sum_{i=1}^{k} \bar{C}_i^{(2)}$	$\bar{C}_k^{(3)}$	$\sum_{i=1}^{k} \bar{C}_i^{(3)}$	$\bar{C}_k^{(4)}$	$\sum_{i=1}^{k} \bar{C}_i^{(4)}$
1	1.72	——	1.84	——	2.07	——	8.98	——
2	1.59	3.31	1.63	3.47	1.65	3.72	4.61	13.56
3	1.07	4.38	1.02	4.49	1.04	4.76	0.62	14.18
4	0.87	5.25	0.70	5.19	0.59	5.35	0.23	14.41
5	0.75	6.00	0.57	5.76	0.41	5.76	0.13	14.54
6	0.66	6.66	0.44	6.20	0.35	6.11	0.13	14.67
7	0.51	7.17	0.35	6.55	0.30	6.41	0.12	14.79
8	0.46	7.63	0.30	6.85	0.27	6.68	0.12	14.91
9	0.42	8.05	0.29	7.14	0.27	6.95	0.13	15.04
10	0.38	8.43	0.29	7.43	0.27	7.22	0.13	15.17
Sum =	8.43		7.43		7.22		15.17	

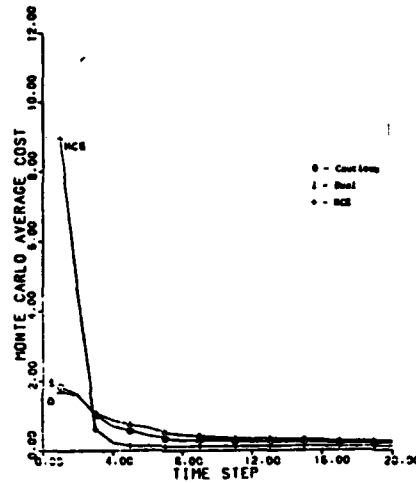Table II. Average costs for the four algorithms in the Example.

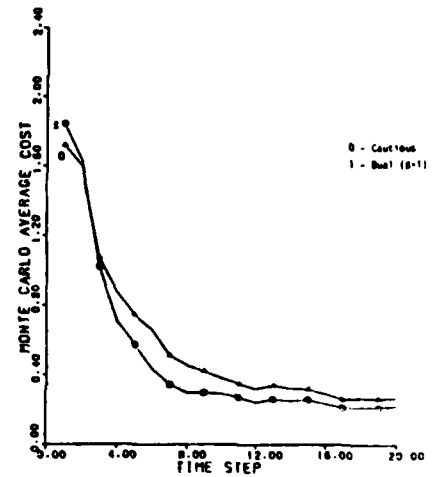Figure 1. Comparison of the performance of the cautious, dual, and HCE controllers

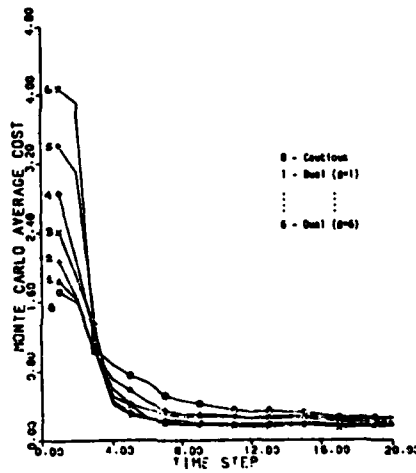Figure 3. Comparison of the performances of cautious and dual algorithms

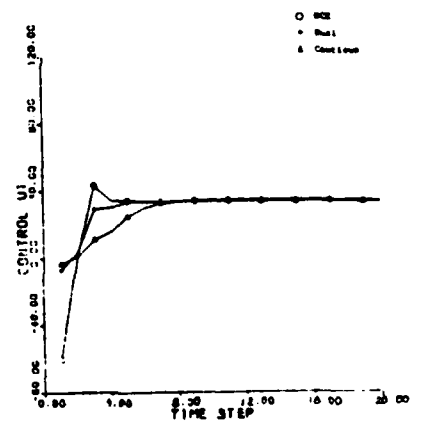Figure 2. Comparison of the performances of cautious, dual and modified dual algorithms

Figure 4. Comparison of the Time History of Control with Cautious, Dual and HCE Controllers
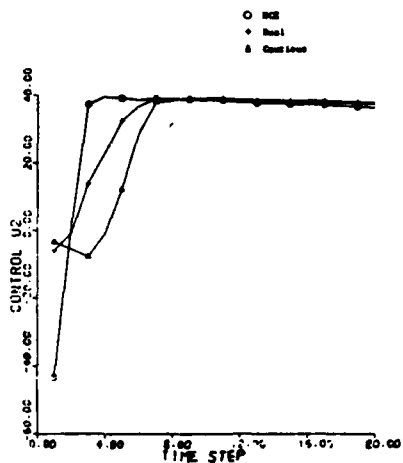
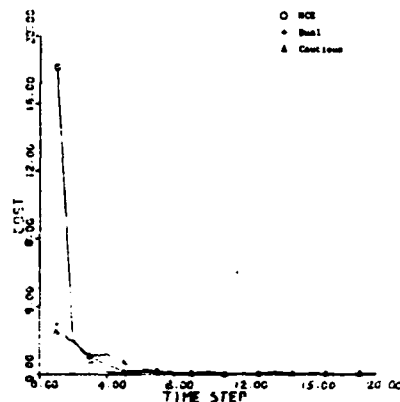Figure 5. Comparison of the Time History of Control U2 with Cautious, Dual and HCE Controllers

Figure 6. Comparison of the Time History of Cost with Cautious, Dual and HCE Controllers

7. Conclusion

A suitable expansion of the cost to go in the stochastic dynamic programming equation can yield a linear controller that accounts for the controller's dual effect.

The simulation runs indicate that a dual controller under certain situations shows up to 49% improvement over the HCE and cautious control-lers. Statistical analysis of Monte Carlo runs indicates that on the average use of the dual controller provides approximately a 20% improvement in the performance criteria over the cautious controller.

For the HCE controller the learning of the parameters is faster than the dual or cautious but the vibration cost is more. As a trade-off be-tween faster convergence and lesser cost, the dual controller seems to be the best.

Appendix A

Statistical Significance in the Comparison of Controller Performance

Two control algorithms are compared by performing a Monte Carlo simu-lation. S independent runs with the two algorithms, under the same homo-geneous conditions, yield a set of i.i.d. samples $C_{ik}^{(1)}$, $C_{ik}^{(2)}$, $i=1,2,\ldots,S$ from two distributions with true but unknown means $J_k^{(1)}$ and $J_k^{(2)}$, respectively, for each time step k.

The sample means

$$\bar{c}_k^{(j)} = \frac{1}{S} \sum_{i=1}^{S} c_{ik}^{(j)} \qquad j=1,2 \tag{A.1}$$

are point estimates of the respective true means.

A statement that

$$\bar{c}_k^{(1)} < \bar{c}_k^{(2)} \tag{A.2}$$

indicating that algorithm 1 is better than 2 for time step k has to be accompanied by a level of significance α of type I error.

Thus we test the hypothesis

$$H_0: \quad \Delta = J_k^{(2)} - J_k^{(1)} \le 0 \qquad \text{(algorithm 1 not better)} \tag{A.3}$$

against the one sided alternative

$$H_1: \quad \Delta = J_k^{(2)} - J_k^{(1)} > 0 \qquad \text{(algorithm 1 better)} \tag{A.4}$$

for a particular α level at each time step k.

This probability of error α is defined as

$$\alpha \triangleq P\{\text{accept } H_1/H_0 \text{ true}\} \tag{A.5}$$

Since we get a set of data of the performances of the two algorithms on the plant under similar conditions we regard it as a set of naturally paired observations.

We consider the sample differences

$$\Delta_{ik} = c_{ik}^{(2)} - c_{ik}^{(1)} \tag{A.6}$$

and this set of differences Δ_{ik} represents a sample with mean

$$\Delta_k = J_k^{(2)} - J_k^{(1)} \tag{A.7}$$

Thus we have reduced the two-sample problem to a one-sample problem. The hypothesis is tested by examining whether Δ_k can be accepted as being positive with high confidence. The test statistic is

$$z_k \triangleq \frac{\bar{\Delta}_k}{\sigma_{\bar{\Delta}_k}} \tag{A.8}$$

where

$$\bar{\Delta}_k = \frac{1}{S} \sum_{i=1}^{S} \Delta_{ik} \tag{A.9}$$

$$\sigma_{\bar{\Delta}_k}^2 = \frac{1}{S(S-1)} \sum_{i=1}^{S} (\Delta_{ik} - \bar{\Delta}_k)^2 \tag{A.10}$$

The test statistic Z_k has a t – distribution with (S-1) degrees of freedom. For S large (>50) Z has a normal distribution. Then we have

$$\sigma_{\bar{\Delta}_k}^2 \approx \frac{1}{S^2} \sum_{i=1}^{S} (\Delta_{ik} - \bar{\Delta}_k)^2 \tag{A.11}$$

and the hypothesis H_1 is accepted if

$$Z_k > c \tag{A.12}$$

where c is taken from the normal distribution tables. For a 1 sided-test with $\alpha = 0.05$, c = 1.645.

The estimated improvement for each time step k is defined as

$$EI_k \triangleq \frac{\bar{c}_k^{(2)} - \bar{c}_k^{(1)}}{\bar{c}_k^{(2)}} \times 100\% \tag{A.13}$$

REFERENCES

[A1] K. J. Åström, U. Borisson, L. Ljung and B. Wittenmark, "Theory and Application of Self-Tuning Regulators," Automatica, vol. 13, pp. 457-476, Sept. 1977.

[A2] J. Alster and P. Belanger, "A Technique for Dual Adaptive Control," Automatica, vol. 10, pp. 627-634, 1974.

[B1] Y. Bar-Shalom and E. Tse, "Dual Effect, Certainty Equivalence and Separation in Stochastic Control, "IEEE Trans. Auto. Control, vol. AC-19, pp. 494-500, Oct. 1974.

[B2] Y. Bar-Shalom and E. Tse, "Concepts and Methods in Stochastic Control", in Control and Dynamic Systems: Advances in Theory and Applications, C. T. Leondes, (ed.), Academic Press, 1976.

[F1] A. A. Feldbaum, Optimal Control Systems, New York: Academic Press, 1965.

[M1] R. Milito, C. S. Padilla, R. A. Padilla and D. Cadorin, "Dual Control Through Innovations," Proc. 19th IEEE Conf. in Decision and Control, Albuquerque, NM, Dec. 1980.

[M2] J. A. Molusis et al., "A Unified Approach to the Optimal Design of Adaptive and Gain Scheduled Controllers to Achieve Minimum Helicopter Rotor Vibration", Proc. 37th Annual American Helicopter Society Conf., New Orleans, LA, May 1981 (to appear in American Helicopter Society Journal, 1983).

[N1] G. E. Noether, Introduction to Statistics, Boston: Houghton Mifflin Company, 1971.

[S1] J. Sternby, "A Regulator for Time-Varying Stochastic Systems,"Proc. 7th IFAC World Congress, pp. 2025-2032, Helsinki, Finland, June 1978.

[W1] C. J. Wenk and Y. Bar-Shalom, "A Multiple Model Adaptive Dual Con-
      trol Algorithm for Stochastic Systems with Unknown Parameters, "IEEE
      Trans. Auto. Control, vol. 25, pp. 703-710, Aug. 1980.

[W2] B. Wittenmark, "An Active Suboptimal Dual Controller for Systems with
      Stochastic Parameters," Automatic Control Theory and Applications
      (Canada), vol. 3, pp. 13-19, 1975.

[W3] A. Weisbrich et al., "Design Study of a Feedback Control System for
      the Multicyclic Flap System Rotor", NASA CR-151960, Jan. 1977.

Acknowledgement

MODEL ADAPTIVE DUAL CONTROL OF MIMO STOCHASTIC SYSTEMS

Carl J. Wenk and Yaakov Bar-Shalom

Department of Electrical Engineering
and Computer Science U-157
University of Connecticut
Storrs, Connecticut 06268

An adaptive dual control algorithm is presented for multiple-input, multiple output (MIMO) linear systems with input and output noise and unknown parameters. The system parameters are assumed to belong to a finite set on which a prior probability distribution is available. The difficulties in characterizing the future evolution if MIMO system information as required by the dynamic programming are overcome through a novel way of using preposterior analysis. This provides a probabilistic characterization of the future adaptation process and allows the controller to take advantage of the dual effect.

1. Introduction

In the control of linear stochastic systems with known dynamics and quadratic cost the Certainty Equivalence (CE) property [A1, B1] is known to hold. When the dynamics are incompletely known, however, due to parameter and noise covariance uncertainty in the system to be controlled, then the CE property does not hold and the dynamic programming cannot be solved [A1]. As shown in [B2] the optimum control has the dual effect: it affects not just the future state of the system, but also the future state, parameter, and noise covariance uncertainty.

To circumvent this inability to compute the optimum solution, a number of adaptive suboptimum control strategies have been developed [S1, D1, A2, T1, W1]. Except for [T1, W1], however, most of these strategies are only passively adaptive [B1]; they do not use the knowledge that future learning will occur. An algorithm using such knowledge to improve its control decisions is called actively adaptive; the dual effect of the control is used to enhance the estimation and identification and ultimately the performance.

This paper presents an actively adaptive control algorithm for multiple-input, multiple-output (MIMO) linear stochastic systems where there is uncertainty in the measurements made on the system, and where the vector Θ of constant but unknown system parameters and noise covariances is equal to one of M known model vectors Θ_j, j=1,...,M. The problem of control of multiple model dynamic systems considered here is a significant generalization of the well known "two-armed bandit problem".

The aspects which make the problem considered here quite general are the in-

clusion of dynamics with discrete uncertainties as well as continuous input and output noises. The algorithm extends the method presented in [W1], which was developed only for single-input, single-output stochastic systems where there was parameter uncertainty but only white input noise with known covariance. The algorithm for MIMO systems in general state-space form presented here is a more sophisticated suboptimal solution to the dynamic programming equation for the multiple model problem than the state-of-the-art algorithms [D1, S1, A2].

The algorithm presented in this paper, called the MIMO Model Adaptive Dual (MAD) control algorithm, overcomes the special difficulties posed by the MIMO system in characterizing the future evolution of information through a novel use of preposterior analysis. Approximate prior probability densities are obtained and used to characterize future learning. The result is an approximate solution to the stochastic dynamic programming, the exact solution to which would give the globally optimum (dual) control.

2. Problem Formulation

Consider controlling a MIMO linear stochastic system whose dynamics and measurements depend on an unknown vector Θ. The system state propagates in discrete time as:

$$x(k+1) = A(\Theta)x(k) + B(\Theta)u(k) + D(\Theta)w(k) \qquad (2.1)$$

where $x(k)$ is the state n-vector, $u(k)$ is the control r-vector, and $w(k)$ is a disturbance d-vector assumed zero mean, white, and Gaussian with variance $W(\Theta)$. Imperfect system measurements are made as:

$$y(k) = H(\Theta) x(k) + v(k) \qquad (2.2)$$

where $y(k)$ is the measurement q-vector and $v(k)$ represents the measurement uncertainty, also taken as zero mean, white, and Gaussian with variance $V(\Theta)$. $w(k)$ and $v(k)$ are assumed uncorrelated. The system matrices $A(\Theta)$, $B(\Theta)$, $D(\Theta)$, $H(\Theta)$, and noise covariances $W(\Theta)$, $V(\Theta)$ are known functions of the constant but unknown vector Θ, which is assumed equal to one of M known constant model vectors Θ_j, j=1, ...,M, with corresponding known a priori probabilities:

$$P[\Theta=\Theta_j] = \Lambda_j(0); \qquad j=1,...,M \qquad (2.3)$$

$$\sum_{j=1}^{M} \Lambda_j(0) = 1 \qquad (2.4)$$

821

The objective is to obtain a control sequence $\{u(0),\ldots,u(N-1)\}$ minimizing

$$J(0) = E[C(0)] \qquad (2.5)$$

where the cost is quadratic about a given, time-varying reference trajectory:

$$C(k) = \frac{1}{2}[x(N) - x_r(N)]' Q(N)[x(N) - x_r(N)]$$

$$+ \frac{1}{2}\sum_{i=k}^{N-1}\{[x(i) - x_r(i)]' Q(i)[x(i) - x_r(i)]$$

$$+ [u(i) - u_r(i)]' R(i) [u(i) - u_r(i)]\} \quad (2.6)$$

subject to equations (2.1)-(2.4). The information vector at time k, Z(k), consists of the measurements and controls up to k:

$$Z(k) = \{y(0),y(1),\ldots,y(k),u(0),u(1),\ldots,u(k-1)\}$$

$$(2.7)$$

The optimum control $u^*(k)$, a function of $Z(k)$ and the statistical description of the future measurements [B1], is obtained by solution of the stocastic dynamic programming:

$$J^*(k) = \min_{u(k)} E\{\frac{1}{2}[x(k) - x_r(k)]' Q(k)[x(k) - x_r(k)]$$

$$+ \frac{1}{2}[u(k) - u_r(k)]' R(k) [u(k) - u_r(k)]$$

$$+ J^*(k+1)|Z(k), u(k)\}. \qquad (2.8)$$

The exact solution of (2.8) is impossible due to the "curse-of-dimensionality" ; the parameter and noise covariance uncertainty prevent the exact computability of $E[J^*(k+1)|Z(k), u(k)]$. The state-of-the-art in suboptimum algorithms which circumvent this difficulty has largely consisted of the Heuristic Certainty Equivalence (HCE) algorithm [B1], where

$$\hat{\theta}(k) = \sum_{j=1}^{M} \Lambda_j(k)\theta_j \qquad (2.9)$$

is assumed the true parameter vector, and the Deshpande-Upadhyay-Lainiotis (DUL) algorithm [D1], where the model-optimal controls $u_j(k)$ are computed and the actual control taken as

$$u(k) = \sum_{j=1}^{M} \Lambda_j(k) u_j(k) \qquad (2.10)$$

The active Model Adaptive Dual control algorithm (MAD) developed in [W1] for systems in input-output form was able to achieve significant performance superiority over the passively adaptive (non-dual) HCE and DUL algorithms by directly obtaining an accurate approximation of $E[J^*(k+1)|Z(k), u(k)]$.

3. **Approximate Solution of the Stochastic Dynamic Programming Equation by Pairwise Preposterior Model Discrimination**

The computation of $E[J^*(k+1)|Z(k), u(k)]$ in

the solution of the stochastic dynamic programming equation (2.8) for M models can be reduced to computing $M(M-1)/2$ two-model costs by use of a result which may be found in [W1]. Only the two-model cost approximation will be developed here using models θ_1, θ_2. The prior probabilities at k in the two-model problem are

$$P[\theta=\theta_1|Z(k), u(k)] = \Pi(k), \quad P[\theta=\theta_2|Z(k),u(k)]$$

$$= 1-\Pi(k) \qquad (3.1)$$

For computational feasibility the cost is approximated as follows: the future controls $(i \geq k+1)$ are assumed to be of the DUL type structure with time-varying probabilities as more information becomes available to the controller. Thus

$$E[J^*(k+1)|Z(k), u(k)] \cong$$

$$E\{\min_{L(k+1)} E[C(k+1)|Z(k+1),L(k+1)]|Z(k),u(k)\} \quad (3.2)$$

where $L(k+1)$ is the set of parameters in the controller structure from k+1 through the end. Using the total probability theorem the (approximation of the) optimum cost-to-go may be written as

$$J^*(k+1) \cong \min_{L(k+1)} \{\Pi(k+1)E[C(k+1)|Z(k+1),L(k+1),\theta=\theta_1]$$

$$+ [1-\Pi(k+1)]E[C(k+1)|Z(k+1),L(K+1),$$

$$\theta=\theta_2]\} \qquad (3.3)$$

where by Bayes' rule

$$\Pi(k+1) = P[\theta=\theta_1|Z(k+1)]$$

$$= \left[1 + \frac{1-\Pi(k)}{\Pi(k)} \frac{p[y(k+1)|Z(k),u(k),\theta=\theta_2]}{p[y(k+1)|Z(k),u(k),\theta=\theta_1]}\right]^{-1} \quad (3.4)$$

with the appropriate Gaussian densities in (3.4) being

$$p[y(k+1)|Z(k), u(k), \theta=\theta_j] =$$

$$N[y(k+1); \hat{y}_j(k+1)|k), S_j(k+1)|k)] \qquad (3.5)$$

where the means and variances in (3.5) are obtained from two Kalman filters, matched to $\theta=\theta_j$, $j=1,2$, respectively.

Next note that (3.2) requires performing a multiple integration over the elements of $y(k+1)$. This is not computationally feasible, in general, and will be avoided through the following procedure. From (3.4) and (3.5) it can be seen that $y(k+1)$ and $\Pi(k+1)$ are related through a mapping described by

$$\frac{1}{2}(y-\hat{y}_1)'S_1^{-1}(y-\hat{y}_1) - \frac{1}{2}(y-\hat{y}_2)' S_2^{-1}(y-\hat{y}_2) =$$

$$\ln\left[\frac{|S_2|^{q/2}}{|S_1|^{q/2}} \frac{\Pi(k)[1-\Pi(k+1)]}{[1-\Pi(k)]\Pi(k+1)]}\right] \qquad (3.6)$$

where the time arguments of $y(k+1)$, $\hat{y}_j(k+1)|k)$ and $S_j(k+1)|k)$ have been dropped. Since a given

822

$\Pi(k+1)$ may result from an infinite number of $y(k+1)$, it is clear that $\Pi(k+1)$ is not a sufficient statistic for $y(k+1)$. However, $\Pi(k+1)$ can be used to serve as an "approximate sufficient statistic". Thus (3.3) may be rewritten as

$$J^\star(k+1) \cong \min_{L(k+1)} \{\Pi(k+1)E[C(k+1)|Z(k),u(k),$$

$$\Pi(k+1),\ L(k+1),\ \theta=\theta_1]$$

$$+ [1-\Pi(k+1)]E[C(k+1)|Z(k),u(k),\Pi(k+1),$$

$$L(k+1),\ \theta=\theta_2]\} \qquad (3.7)$$

The outer expectation of (3.2) over $y(k+1)$ is then replaced by an expectation with respect to $p[\Pi(k+1)|Z(k),\ u(k)]$, the preposterior probability density [R1] of $\Pi(k+1)$, the "model information state" at k+1. An approximate preposterior density with two delta functions at locations $\Pi_1(k+1)$ and $\Pi_2(k+1)$ is used as in [B3,W1].

Having established an implementable preposterior density, the next step is to construct the minimization in (3.7) with respect to the time-varying future controller parameter set $L(k+1)$, a set depending of course on $\Pi(k+1)$. An easily implemented approximate solution to this minimization is obtained by assuming a future sequence of DUL controls represented by $L(k+1)$:

$$E[J^\star(k+1)|Z(k),\ u(k)] \sim \hat{J}(k+1)$$

$$= \int_0^1 \{\Pi(k+1)E[C(k+1)|Z(k),u(k),\Pi(k+1),\bar{L}(k+1),\theta=\theta_1]$$

$$+ [1-\Pi(k+1)]E[C(k+1)|Z(k),u(k),\Pi(k+1),\bar{L}(k+1),\theta=\theta_2]\}$$

$$\cdot p[\Pi(k+1)|Z(k),\ u(k)]\ d\Pi(k+1) \qquad (3.8)$$

Using the two delta function preposterior density above and performing the integration gives the approximate cost-to-go resulting from a particular control decision $u(k)$:

$$\hat{J}(k+1) = \Pi(k)\Pi_1(k+1)\hat{J}_{11}[k+1),u(k),\bar{L}_{11}(k+1),\theta=\theta_1]$$

$$+ \Pi(k)[1-\Pi_1(k+1)]\hat{J}_{12}[k+1,\ u(k),\bar{L}_{12}(k+1),\theta=\theta_2]$$

$$+ [1-\Pi(k)]\Pi_2(k+1)\hat{J}_{21}[k+1,u(k),\bar{L}_{21}(k+1),\theta=\theta_1]$$

$$+ [1-\Pi(k)][1-\Pi_2(k+1)]\hat{J}_{22}[k+1,u(k),\bar{L}_{22}(k+1),\theta=\theta_2] \quad (3.9)$$

The nominal sequence of control parameters $\bar{L}_{\ell j}(k+1)$, ℓ, j=1,2 comes from a time-varying DUL weighted sum of model-optimal controls. This sum is computed with nominal weighting factors given by:

(i) $\Pi(k+1) = \Pi_\ell(k+1)$ as the sufficient statistic for θ at k+1,

(ii) subsequent nominal posterior probabilities $\Pi_{\ell j}(i)$ that $\theta=\theta_1$ which evolve as i=k+2,..., N-1 when this DUL control is applied to the system with $\theta=\theta_j$.

The single-model optimal control parameters are obtained from a standard linear quadratic problem with θ known. The costs $\hat{J}_{\ell j}$ are obtained

from a recursion for the linear system with $\theta=\theta_j$, quadratic cost, using a DUL control policy with control parameters $\bar{L}_{\ell j}(k+1)$. Details of the nominal posterior probability generation and the recursions for $\hat{J}_{\ell j}$ are contained in [W2].

4. Numerical Studies

A second order system is considered with the following two-model system description.

$$A(\theta_1) = A(\theta_2) = \begin{bmatrix} 0 & -0.49 \\ 1 & 1.4 \end{bmatrix}$$

$$B(\theta_1) = [0.45\ \ 2]'$$

$$B(\theta_2) = [0.9\ \ 1]'$$

$$D(\theta_1) = D(\theta_2) = diag(1,1)$$

$$H(\theta_1) = H(\theta_2) = diag(1,1)$$

$$W(\theta_1) = W(\theta_2) = diag(10^{-4},2.25)$$

$$V(\theta_1) = V(\theta_2) = diag(10,10^{-2})$$

A priori, $P(\theta=\theta_1)=P(\theta=\theta_2) = 0.5$. The control objective is to take the initial state of $x(0) = [0\ \ 0.1]'$ and make it follow over N = 5 time stages the state reference trajectory

$$x_r(1) = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix},\ \ x_r(2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ \ x_r(3) = \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

$$x_r(4) = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix},\ \ \ x_r(5) = \begin{bmatrix} 0 \\ 10 \end{bmatrix}.$$

with quadratic weighting matrices

$$Q(0) = 0 \qquad \text{and}$$

$$Q(1) = Q(2) = Q(3) = diag(0,1)$$

$$Q(4) = diag(0,5)\ \ ,\ \ \ Q(5) = diag(0,50)$$

There was no penalty associated with the control, $R(k) = 0\ \forall\ k$.

The first test was to compute the sample means and sample standard deviations of the cost samples C_i^{OPT}, C_i^{HCE}, C_i^{DUL}, and C_i^{MAD}. The results are contained in Table 1.

Algorithm	OPT	HCE	DUL	MAD
Sample mean	60.97	269.3	223.4	110.4
Sample standard deviation	73.9	443.6	406.4	137.8

Table 1. Sample Average Costs and Standard Deviations

This table gives the first indication of the superiority of MAD over HCE and DUL in both mean cost reduction and performance cost variability. Note that MAD has reduced the mean cost by 51% over DUL, and by 59% over HCE. MAD has reduced the cost variability by 66% over DUL and by 69% over HCE.

Are these results truly statistically signifi-

cant? Are the true means ordered as the sample means would indicate? To answer these questions, a rigorous statistical test for the comparison of controller performances was developed in [W1]. The sample means of the differences and the standard deviations of the sample means are given for the algorithms in Table 2. They indicate that

Algorithms Compared	$\bar{\Delta}$	σ_Δ	Significance test statistic $\Delta/\sigma_{\bar{\Delta}}$	Estimated Improvement %
HCE - DUL	45.866	13.767	3.3316	17
HCE - MAD	158.86	29.881	5.3164	59
DUL - MAD	112.99	27.033	4.1797	51

Table 2. Statistical test results for algorithm comparisons

the hypotheses that MAD is better than both HCE and DUL are accepted.

Table 3 illustrates the manner in which the need for active learning is sensed by MAD. For various possible values of the control decision at period 1, MAD evaluates the future learning opportunities and calculates the future costs. For $u(1) = 4.35$ the preposterior density characterized by $\Pi_1(2)$ and $\Pi_2(2)$ indicates that not enough learning will take place to minimize the effect of the term $J_{21}(2)$ in the cost to go equation (3.9). ($J_{21}(2)$ represents the cost of a mismatched controller which does not learn fast enough what the true system is). For larger $u(1)$ the learning is faster, but after $u(1) = 5.09$ the price of learning exceeds the benefit.

Table 3 also gives insight into how to determine a priori (or even on line), in a non-Monte Carlo fashion, when it is valuable (and necessary) to use an active, dual control decision making algorithm like MAD: when the penalty for a mismatched controller is large and its probabilistic contribution to the cost is significant.

	$u(1)$	$\hat{J}(2)$	$\Pi_1(2)$	$\Pi_2(2)$	$\hat{J}_{11}(2)$	$\hat{J}_{12}(2)$	$\hat{J}_{21}(2)$	$\hat{J}_{22}(2)$
	4.35	126.5	.9416	.05747	64.93	68.49	2153.	66.97
	4.55	104.0	.9488	.05039	65.68	68.04	1557.	66.45
	4.75	76.00	.9554	.04389	66.59	67.80	504.5	65.98
MAD	5.09	67.76	.9653	.03415	68.56	67.66	116.2	65.25
	5.35	68.19	.9713	.02824	70.29	67.75	114.4	64.79
	5.55	68.78	.9754	.02421	71.85	67.93	121.5	64.48
	5.75	69.47	.9790	.02067	73.56	68.20	129.8	64.20

Table 3. Cost Breakdown and Learning for MAD

5. Concluding Remarks

An actively adaptive control algorithm has been derived for multiple input, multiple output stochastic systems in general state space form possessing both continuous and discrete modes of system uncertainty. The algorithm, called Model Adaptive Dual Control, is the only actively adaptive controller for this class of systems. Rigorous statistical tests were used to show statistically significant performance improvement in the new actively adaptive MIMO MAD algorithm over two state-of-the-art passively adaptive control algorithms. It has been shown in particular that when there is heavy terminal state penalty and the control period is relatively short, passive learning often does not suffice.

In these cases active adaptation can be expected to improve the transient behavior in adaptive control by speeding up the adaptation process.

References

[A1] M. Aoki, Optimization of Stochastic Systems, New York: Academic Press, 1967.

[A2] M. Athans, et.al.,"The Stochastic Control of the F-8C Aircraft Using a Multiple Model Adaptive Control Method, IEEE Tr. Auto. Control, AC-22, pp. 768-780, Oct., 1977.

[B1] Y. Bar-Shalom and E. Tse, "Concepts and Methods in Stochastic Control", in C. T. Leondes (Ed.) Control and Dynamic Systems: Advances in Theory and Applications, Vol. 12, Academic Press, 1976.

[B2] Y. Bar-Shalom and E. Tse, "Dual Effect, Certainty Equivalence, and Separation in Stochastic Control", IEEE Tr. Auto. Control, Vol. AC-19, pp. 494-500, Oct., 1974.

[B3] Y. Bar-Shalom, R. E. Larson, M. A. Grossberg, "Application of Stochastic Control Theory to Resource Allocation Under Uncertainty", IEEE Tr. Auto. Control, Vol. AC-19, pp. 1-7, Feb., 1974.

[D1] J. G. Deshpande, T. N. Upadhyay, and D. G. Lainiotis, "Adaptive Control of Linear Stochastic Systems", Automatica, Vol. 9, pp 107-115, 1973.

[S1] G. N. Saridis, Self-Organizing Control of Stochastic Systems, Dekker, 1977.

[T1] E. Tse and Y. Bar-Shalom, "An Actively Adaptive Control for Discrete-Time Systems with Random Parameters", IEEE Tr. Auto. Control, AC-18, pp. 109-117, April, 1973.

[W1] C. J. Wenk and Y. Bar-Shalom, "A Multiple Model Adaptive Dual Control Algorithm for Stochastic Systems with Unknown Parameters", IEEE Tr. Auto. Control, AC-25, No. 4, pp. 703-710, Aug., 1980.

[W2] C. J. Wenk, "Decision Strategies with Dual Effect for Inseparable Stochastic Control Problems with Continuous and Discrete Uncertainty", Ph. D. Thesis, Univ. of Conn., EECS Dept., April,1981.

Dual Control Guidance for Simultaneous Identification and Interception*

K. BIRMIWAL† and Y. BAR-SHALOM†

An adaptive dual-control guidance algorithm enables moving target interception in the presence of an interfering target when noisy, nonlinear, state dependent feature measurements are available for target identification.

Key Words—Guidance systems; dual control; Kalman filters; identification; dynamic programming.

Abstract—An adaptive dual-control guidance algorithm is presented for intercepting a moving target in the presence of an interfering target (decoy) in a stochastic environment. Two sequences of measurements are obtained at discrete points in time; however, it is not certain which sequence came from the target of interest and which from the decoy. Associated with each track, the interceptor also receives noisy, state-dependent feature measurements. The optimum control for the interceptor which is given by the solution of the stochastic dynamic programming equation is not numerically feasible to obtain. An approximate solution of this equation is obtained by evaluating the value of the future information gathering. This is done through the use of preposterior analysis: approximate prior probability densities are obtained and used to describe the future learning and control. In this way, the interceptor control is used for information gathering in order to reduce the future target and decoy inertial measurement errors and enhance the observable target decoy feature differences for subsequent discrimination between the true target and the decoy. Simulation studies have shown the effectiveness of the scheme.

1. INTRODUCTION

A NEW CONTROL-DECISION strategy for intercepting a moving target is developed where the target is using a defensive decoy in an environment best described by a stochastic process. The decision-making problem takes place during the terminal phase of interceptor guidance.

At discrete points in time the interceptor receives noisy, state-dependent, feature measurements: one from the true target and one from the decoy. It is assumed that there is no measurement to track association uncertainty; however, it is not certain which measurement sequence came from the target and which from the decoy. Additional sources of uncertainty are the imperfect, noise-corrupted state

observations and the inherently unknown time-to-intercept. The result is a highly nonlinear stochastic control and decision-making problem, with both continuous (all noises) and discrete (track identity) sources of uncertainty, in which the control has a dual effect (Feldbaum, 1965): in addition to its effect on the relative interceptor/target/decoy states themselves, the present interceptor control also affects the future feature observation process and hence the target decoy identification uncertainty. Specifically, the interceptor control must be used for information gathering about the true target track by: (a) reducing future target and decoy inertial measurement errors by changing its own state and hence the relative states, and by (b) enhancing observable target decoy feature differences for subsequent discrimination between the true target and the decoy. All of these information theoretic characteristics are functions of the interceptor/target decoy states, which are in turn directly affected by the interceptor control. The decisions must also simultaneously be used to optimize the function of interceptor guidance toward the target (control proper, which is inseparable from the information gathering). The problem is further complicated by certain constraints: maximum fuel capability, and possibly, maximum time-to-intercept and interceptor state constraints.

This is an example of a nonlinear stochastic control problem in which the optimum solution exhibits an inseparability between the dual actions of the control decision in gathering information about the partially unknown system (reducing uncertainty), and simultaneously changing the system state itself (the control function proper, which requires minimum uncertainty or maximum information about the state). In general, systems with both continuous and discrete nonlinear probabilistic structures create decision-making

Gaussian, zero mean (WGZM) with known co-variance Q_I.

Let the motion of the interceptor be given by

$$\mathbf{x}_1(k + 1) = A_1\mathbf{x}_1(k) + B\mathbf{u}(k) + G_1\mathbf{w}_1(k)$$

$$k = 0, 1, 2, \ldots \qquad (2)$$

where $\mathbf{u}(k)$ is the interceptor control vector to be determined at time k and $\mathbf{w}_1(k)$ is the process noise, WGZM with known covariance Q_1.

The measurement equations of the two vehicles are

$$\mathbf{z}_I(k) = H_I\mathbf{x}_I(k) + \mathbf{v}_I(k) \quad I = 1, 2 \qquad (3)$$

$$k = 1, 2, \ldots$$

where $\mathbf{v}_I(k)$ is the measurement noise, WGZM with known covariance R_I.

The measurement equation associated with the interceptor is

$$\mathbf{z}_1(k) = H_1\mathbf{x}_1(k) + \mathbf{v}_1(k) \quad k = 1, 2, \ldots \qquad (4)$$

where $\mathbf{v}_1(k)$, the measurement noise of the interceptor, is assumed to be WGZM with known covariance R_1.

To discriminate between the two vehicles, a feature measurement $\beta_I(k)$ associated with each vehicle I is obtained. For simplicity, this feature measurement is assumed to be a scalar and is a function of the state of the vehicle $\mathbf{x}_I(k)$, the state of the interceptor $\mathbf{x}_1(k)$ and the true feature ϕ_I, that is

$$\beta_I(k) = f[\phi_I, \mathbf{x}_I(k), \mathbf{x}_1(k)] + \chi_I(k) \quad I = 1, 2, \qquad (5)$$

$$k = 1, 2, \ldots$$

Here, it is assumed that $\phi_1 \neq \phi_2$ for the identification purpose and $\chi_I(k)$ is the additive white noise, independent of the states, assumed normal with mean zero and variance σ_I^2.

All the noise sequences are assumed to be mutually independent.

In this formulation of the target decoy interception problem, it is assumed that the vehicles follow the state equations (1) without changing their state models. The extension of this formulation to the case of the target decoy changing its state model is discussed in the example section. Also, this model can easily be extended to the case of state-dependent feature measurement noise.

The following notations are used:

$$Z^k = \{\mathbf{z}_I(i), \mathbf{z}_1(i): I = 1, 2; i = 1, 2, \ldots, k\} \qquad (6)$$

$$\beta^k = \{\beta_I(i): I = 1, 2; i = 1, 2, \ldots, k\} \qquad (7)$$

$$U^k = \{\mathbf{u}(i): i = 0, 1, \ldots, k\} \qquad (8)$$

$$\pi(k) = P\{\theta = 1 | Z^k, \beta^k, U^{k-1}\} \qquad (9)$$

where $\theta = j, j = 1, 2$ represents the event that the jth track is the track originated from the target of interest.

The interceptor's objective is to choose the control strategy $\mathbf{u}(k)$ that minimizes the expected terminal weighted relative position of the target/interceptor (T/I) at the unknown (random) terminal time N, subject to the dynamic control effort bound and speed limit of the interceptor. For the problem to be meaningful, it is assumed that the interceptor is capable of intercepting any of the vehicles in finite time. This leads to the stochastic control cost criterion to be minimized at time k

$$J(k) = E[C(k)]$$

$$= E\left[\sum_{i=k}^{N-1} \mathbf{u}'(i)R(i)\mathbf{u}(i) + \mathbf{g}'[\mathbf{x}_\theta(N), \mathbf{x}_1(N)] \right.$$

$$\left. \mathbf{Q}\mathbf{g}[\mathbf{x}_\theta(N), \mathbf{x}_1(N)] | Z^k, \beta^k, U^{k-1}\right] \qquad (10)$$

subject to

$$|u_n(i)| \leq u_n^{max}(i) \qquad \forall n, \forall i \geq k \qquad (11)$$

and

$$\tau_1(i + 1) \leq \tau_1^{max} \qquad \forall i \geq k \qquad (12)$$

where $R(i)$ is a known (time-varying) control weighting matrix; $\mathbf{g}[\mathbf{x}_\theta(N), \mathbf{x}_1(N)]$ is a vector-valued function, whose components are the positions differences between the states $\mathbf{x}_\theta(N)$ and $\mathbf{x}_1(N)$; $\mathbf{Q}$ is a known constant weighting matrix associated with this relative terminal T/I position state; $u_n(i)$ is the nth component of the control vector $\mathbf{u}(i)$; $u_n^{max}(i)$ is a known, time-varying dynamic control effort bound, which depends on the kinematic acceleration capability of the interceptor; $\tau_1(i)$, a function of $\mathbf{x}_1(i)$ is the interceptor's speed at time i and τ_1^{max} is a known speed limit of the interceptor.

Let an admissible control decision vector $\mathbf{u}(k)$ be a function of Z^k and β^k as well as the statistical description of the future observations (Bar-Shalom and Tse, 1976). Then the optimum control strategy for this nonlinear stochastic control problem is obtained by applying the Bellman's Principle of Optimality, which leads to the stochastic dynamic programming (SDP) equation. Solution of the SDP equation yields the globally optimal control, which, in general, has the dual effect (Bar-Shalom and Tse, 1974; Feldbaum, 1965). At time k, the SDP is described for this problem as

$$J^*(k) = \min_{\mathbf{u}(k)} E[\mathbf{u}'(k)R(k)\mathbf{u}(k)$$

$$+ J^*(k + 1)|Z^k, \beta^k, U^k] \qquad (13)$$

subject to

$$|u_n(k)| \le u_n^{max}(k) \qquad \forall n \qquad (14)$$

and

$$\tau_1(k + 1) \le \tau_1^{max} \qquad (15)$$

where for a given $u(k)$, $J^*(k + 1)$ is the optimum cost-to-go from time $k + 1$ to the unknown terminal time N and the expectation is done with respect to all future random variables, including both inertial observation errors and the feature parameter observation errors.

The exact solution to this problem is impossible due to the fact that no distribution over N is available and because of the 'curse of dimensionality' (Bellman, 1961). This can be avoided only by a recursion for the cost-to-go which here does not exist because of the track uncertainty. We present next an approximate solution of this problem.

3. APPROXIMATE SOLUTION OF THE STOCHASTIC DYNAMIC PROGRAMMING EQUATION

For computational feasibility, the cost is approximated as follows: the future control $(i \ge k + 1)$ are assumed to be of the DUL type (the 'partitioned' control obtained by Deshpande, Upadhyay and Lainiotis, 1973) as

$$u(i) = \pi(i)u_1(i) + [1 - \pi(i)]u_2(i) \qquad (16)$$

where $u_j(i)$ is the bounded optimum control at time i, given $\theta = j$ with time-varying probabilities as more information becomes available to the controller, and where the controls $u_j(i)$ and $u(i)$ satisfy the constraints (11) and (12) as shown in the next section. With this the optimal cost-to-go in (13) is replaced by

$$E[J^*(k + 1)|Z^k, \beta^k, U^k] \cong$$

$$E\left\{ \min_{L(k+1)} E[C(k + 1)|Z^{k+1}, \beta^{k+1}, U^k, L(k + 1)]\, \Big|\right.$$

$$\left. Z^k, \beta^k, U^k \right\} \qquad (17)$$

where $L(k + 1)$ is the set of parameters in the controller structure from $k + 1$ through the end and $C(k + 1)$ is the cost function. Using the total probability theorem, the (approximation of the) optimum cost-to-go may be written as

$$J^*(k + 1) \cong \min_{L(k+1)} \{\pi(k + 1)E[C(k + 1)|Z^{k+1},$$

$$\beta^{k+1}, U^k, L(k + 1), \theta = 1] + [1 - \pi(k + 1)]$$

$$\cdot E[C(k + 1)|Z^{k+1}, \beta^{k+1}, U^k, L(k + 1), \theta = 2]\}$$

$$(18)$$

where by Bayes' rule

$$\pi(k + 1) = P\{\theta = 1|Z^{k+1}, \beta^{k+1}, U^k\}$$

$$= \left[1 + \left(\frac{1 - \pi(k)}{\pi(k)} \right) \right.$$

$$\left. \frac{p[z(k + 1), \beta(k + 1)|Z^k, \beta^k, U^k, \theta = 2]}{p[z(k + 1), \beta(k + 1)|Z^k, \beta^k, U^k, \theta = 1]} \right]^{-1}$$

$$(19)$$

where

$$z(k + 1) \triangleq [z_1'(k + 1), z_2'(k + 1), z_i'(k + 1)]' \qquad (20)$$

and

$$\beta(k + 1) \triangleq [\beta_1(k + 1), \beta_2(k + 1)]'. \qquad (21)$$

Here $z(k + 1)$ and $\beta(k + 1)$ are respectively the (column) vectors of all state measurements and feature measurements at time $k + 1$.

Assuming that the conditional joint density of $z(k + 1)$, $\beta(k + 1)$ in (19) is known or can be obtained, the computation of (17) requires performing a multiple integration over their elements (20) and (21). This is not computationally feasible and is avoided as follows: since the mapping from $z(k + 1)$ and $\beta(k + 1)$ to $\pi(k + 1)$ is not one-to-one (in fact, many-to-one), $\pi(k + 1)$ is not a sufficient statistic for $z(k + 1)$ and $\beta(k + 1)$. However, $\pi(k + 1)$ can be used to serve as an 'approximate sufficient statistic'. Using this approximate statistic in (18) and then replacing the outer expectation of (17) over $z(k + 1)$ and $\beta(k + 1)$ by an expectation over $\pi(k + 1)$ results in

$$E[J^*(k + 1)|Z^k, \beta^k, U^k] \cong$$

$$\int_0^1 \min_{L(k+1)} \{\pi(k + 1)E[C(k + 1)|Z^k, \beta^k, U^k,$$

$$\pi(k + 1), L(k + 1), \theta = 1]$$

$$+ [1 - \pi(k + 1)]E[C(k + 1)|Z^k, \beta^k, U^k,$$

$$\pi(k + 1), L(k + 1), \theta = 2]\}$$

$$\cdot p[\pi(k + 1)|Z^k, \beta^k, U^k]\, d\pi(k + 1) \qquad (22)$$

where $p[\pi(k + 1)|Z^k, \beta^k, U^k]$ is the preposterior probability density of $\pi(k + 1)$ (Raiffa and Schlaifer, 1972). The use of the exact density in (22) would require numerical integration and this is avoided using a two-point delta function density as in Wenk and Bar-Shalom (1980) and Wenk (1981).

As the vehicles' discrimination capability increases, the preposterior density exhibits a bimodal character, largely concentrated around two distinct locations, say $\pi_1(k + 1)$ and $\pi_2(k + 1)$. The

approximate preposterior density then can be taken as

$$p[\pi(k+1)|Z^k, \beta^k, U^k] \cong \pi(k)\delta[\pi(k+1) - \pi_1(k+1)] + [1 - \pi(k)]$$
$$\delta[\pi(k+1) - \pi_2(k+1)] \quad (23)$$

where the delta function locations $\pi_1(k+1)$ and $\pi_2(k+1)$ satisfy

$$0 \le \pi_2(k+1) \le \pi(k) \le \pi_1(k+1) \le 1. \quad (24)$$

The locations $\pi_1(k+1)$ and $\pi_2(k+1)$ are obtained by matching the first two moments produced by the approximate density (23) to the true preposterior moments of $\pi(k+1)$. The explicit expressions for π_1 and π_2 are derived in the Appendix A (Wenk, 1981). Substituting this simple preposterior density (23) in (22) and assuming that the minimization in (22) occurs when $L(k+1) = \bar{L}(k+1)$ representing the future controls to be of the constrained DUL type, gives approximately the expected cost-to-go resulting from a particular control decision $u(k)$ (Wenk and Bar-Shalom, 1980)

$$E[J^*(k+1)|Z^k, \beta^k, U^k] \cong \pi(k)\pi_1(k+1)\hat{J}_{11}(k+1)$$
$$+ \pi(k)[1 - \pi_1(k+1)] \cdot \hat{J}_{12}(k+1) + [1 - \pi(k)]$$
$$\cdot \pi_2(k+1)\hat{J}_{21}(k+1) + [1 - \pi(k)][1 - \pi_2(k+1)]$$
$$\cdot \hat{J}_{22}(k+1) \quad (25)$$

where

$$\hat{J}_{mj}(k+1) \doteq E[C(k+1)|Z^k, \beta^k, U^k, \bar{L}_{mj}(k+1)]$$
$$\cong \min_{L(k+1)} E[C(k+1)|Z^k, \beta^k, U^k,$$
$$\pi_m(k+1), L(k+1), \theta = j]. \quad (26)$$

The nominal sequence of control parameters $\bar{L}_{mj}(k+1)$; $m, j = 1, 2$ are given by

(i) $\pi(k+1) = \pi_m(k+1)$ as the sufficient statistic for θ at $k+1$, and
(ii) subsequent nominal posterior probabilities $\bar{\pi}_{mi}(i)$ for $i \ge k+2$, representing the probability at time i of the first track being from the target when $\pi(k+1) = \pi_m(k+1)$ and $\theta = j$.

Details of the nominal posterior probability generation and computation of $\hat{J}_{mj}$ are contained in the next section.

4. GENERATION OF THE NOMINAL PARAMETERS AND THE COST-TO-GO

The intercept time N is not necessarily the same for the target and for the decoy. For both the tracks, N is a complicated function of the states of the

vehicle $x_l(k)$ and the interceptor $x_l(k)$, the future controls to be applied and the process noise, yet to come. To obtain a solution of this nonlinear stochastic control problem, N is taken to be the same for both tracks and is estimated as the minimum number of sampling intervals including k in which the interceptor will intercept either of the two vehicles maintaining its control effort bound and its speed limit. Clearly, N is reestimated at each time k and the corresponding estimate is $\hat{N}(k)$.

The nominal sequences of future posterior probabilities $\bar{\pi}_{mj}(i)$; $m, j = 1, 2$ are generated by constructing a future observation and control sequence, based on the statistical information contained in the approximate preposterior density (23), which in turn is a function of the control $u(k)$. At time k, the nominal values for time $k+1$ and for the path mj are obtained as follows

$$\bar{N}_{mj}(k+1) = \hat{N}(k) \quad (27)$$
$$\bar{\pi}_{mj}(k+1) = \pi_m(k+1) \quad (28)$$
$$\bar{x}_{lmj}(k+1) = A_l\hat{x}_l(k|k) \quad l = 1, 2 \quad (29)$$
$$\bar{x}_{lmj}(k+1) = A_l\hat{x}_l(k|k) + Bu(k). \quad (30)$$

The nominal optimal control for the interceptor $\bar{u}_{lmj}(k+1)$, where $\theta = j$, $\pi(k+1) = \pi_m(k+1)$ and the interceptor considers the lth track as the track from the target, is given by the solution of the LQG problem (for the estimated terminal time $\bar{N}_{mj}(k+1)$). In case this optimal control exceeds the bound (11), the appropriate bound is used. Then the nominal DUL control for the interceptor at time $k+1$ is given by

$$\bar{u}_{mj}(k+1) = \bar{\pi}_{mj}(k+1)\bar{u}_{1mj}(k+1)$$
$$+ [1 - \bar{\pi}_{mj}(k+1)] \cdot \bar{u}_{2mj}(k+1). \quad (31)$$

If the resulting nominal speed of the interceptor of time $k+2$ exceeds the limit (12), then the magnitude of this nominal control $\bar{u}_{mj}(k+1)$ is reduced by considering the control $t \cdot \bar{u}_{mj}(k+1)$, $0 \le t \le 1$ (i.e. the direction of the desired nominal control is unchanged) so that the interceptor moves at its speed limit.

Observe that the nominal feature measurements at time $k+1$ are not generated since the information of these features is contained in $\pi_1(k+1)$ and $\pi_2(k+1)$.

For time $i \ge k+2$, the quantities $\bar{x}_{lmj}(i)$, $\bar{x}_{lmj}(i)$, $\bar{\beta}_{mj}(i)$, $\bar{\pi}_{mj}(i)$, $\bar{N}_{mj}(i)$ and $\bar{u}_{mj}(i)$ are obtained recursively as follows:

$$\bar{x}_{lmj}(i) = A_l\bar{x}_{lmj}(i-1) \quad l = 1, 2 \quad (32)$$
$$\bar{x}_{lmj}(i) = A_l\bar{x}_{lmj}(i-1) + B\bar{u}_{mj}(i-1) \quad (33)$$
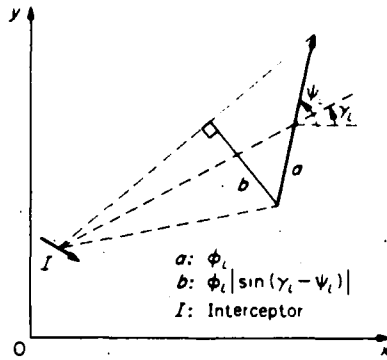$$\bar{\beta}_{lmj}(i) = f[\phi_{lj}, \bar{x}_{lmj}(i), \bar{x}_{lmj}(i)] \quad l = 1, 2 \quad (34)$$

FIG. 1. Projection (b) of the length feature ϕ_l of the lth vehicle (a) on the interceptor's (I) line-of-sight.

$$\beta_l(k) = \phi_l|\sin\,[\gamma_l(k) - \psi_l(k)]| + \alpha_l(k) \quad l = 1, 2$$
$$k = 1, 2, \ldots$$
$$(46)$$

where

$$\gamma_l(k) = \tan^{-1}\frac{y_l(k) - y_1(k)}{x_l(k) - x_1(k)} \qquad -\frac{\pi}{2} \le \gamma_l(k) \le \frac{\pi}{2}$$
$$(47)$$

and

$$\psi_l(k) = \tan^{-1}\frac{\dot{y}_l(k)}{\dot{x}_l(k)} \qquad -\frac{\pi}{2} \le \psi_l(k) \le \frac{\pi}{2}. \qquad (48)$$

Here, though the noise $\alpha_l(k)$ (assumed Gaussian) has the real line as its support, this is an approximation for this example since $\beta_l(k)$ is always non-negative. This approximation is acceptable if $\phi_l \gg \sigma_{\alpha_l(k)}$.

Now, the estimation of the posterior probability $\pi(k + 1)$, when $z(k + 1)$ and $\beta(k + 1)$ are available (i.e. to update the system), the computation of the nominal posterior probability $\tilde{\pi}_{mj}(i)$ (35) the determination of the one-step predicted value of $\beta(k + 1)$, i.e. $\hat{\beta}_j$ (A.4) and the associated covariance S_j (A.7) will complete the discussion for this example. Since the feature measurements (46) are a highly nonlinear function of the states, the conditional joint density of $z(k + 1)$, $\beta(k + 1)$ in (19) is not directly available. Observe that at time k, we do not compute $\pi(k + 1)$ using (19), rather we compute $\pi_1(k + 1)$ and $\pi_2(k + 1)$ using (A.10) and (A.11). Hence, to obtain $\pi(k + 1)$ at time $k + 1$, when the measurements $z(k + 1)$ and $\beta(k + 1)$ are available, we rewrite (19) as

$$\pi(k + 1) = \left[1 + \frac{1 - \pi(k)}{\pi(k)}\right.$$
$$\left.\frac{p[\beta(k + 1)|Z^{k+1}, \beta^k, U^k, \theta = 2]}{p[\beta(k + 1)|Z^{k+1}, \beta^k, U^k, \theta = 1]}\right]^{-1}. \quad (49)$$

Here, if the conditional density of $\beta(k + 1)$ in (49) is determined mainly by the noise characteristics of $\alpha(k + 1)$ (otherwise, a better approximation of this conditional density has to be obtained and this is omitted in this example so as not to deviate from the main theme of this paper), then this density is approximately given by

$$p[\beta(k + 1)|Z^{k+1}, \beta^k, U^k, \theta = j] \cong$$
$$N\left[E[\beta(k + 1)|Z^{k+1}, \beta^k, U^k, \theta = j], \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right]$$
$$j = 1, 2 \quad (50)$$

where

$$E[\beta_l(k + 1)|Z^{k+1}, \beta^k, U^k, \theta = j] \cong$$
$$\phi_{lj}|\sin\,[\hat{\gamma}_l(k + 1|k + 1) - \hat{\psi}_l(k + 1|k + 1)]|$$
$$l = 1, 2. \quad (51)$$

Here, ϕ_{lj} is as in (34) and $\hat{\gamma}_l(k + 1|k + 1)$, $\hat{\psi}_l(k + 1|k + 1)$ are approximated as

$$\hat{\gamma}_l(k + 1|k + 1) \cong$$
$$\tan^{-1}\frac{\hat{y}_l(k + 1|k + 1) - \hat{y}_1(k + 1|k + 1)}{\hat{x}_l(k + 1|k + 1) - \hat{x}_1(k + 1|k + 1)} \quad l = 1, 2$$
$$(52)$$

and

$$\hat{\psi}_l(k + 1|k + 1) \cong \tan^{-1}\frac{\hat{\dot{y}}_l(k + 1|k + 1)}{\hat{\dot{x}}_l(k + 1|k + 1)} \quad (53)$$

with

$$\hat{x}_l(k + 1|k + 1) = E[x_l(k + 1)|Z^{k+1}, U^k] \, l = 1, 2$$
$$(54)$$

$$\hat{x}_1(k + 1|k + 1) = E[x_1(k + 1)|Z^{k+1}, U^k] \quad (55)$$

being obtained using Kalman filters.

Similarly, rewriting (35) to obtain the nominal posterior probability $\tilde{\pi}_{mj}(i)$ gives

$$\tilde{\pi}_{mj}(i) = \left[1 + \frac{1 - \tilde{\pi}_{mj}(i - 1)}{\tilde{\pi}_{mj}(i - 1)}\right.$$
$$\left.\cdot\left\{\frac{p[\bar{\beta}_{mj}(i)|Z^i_{mj}, \beta^{i-1}_{mj}, U^{i-1}_{mj}, \theta = 2]}{p[\bar{\beta}_{mj}(i)|Z^i_{mj}, \beta^{i-1}_{mj}, U^{i-1}_{mj}, \theta = 1]}\right\}\right]^{-1}. \quad (56)$$

As in (49), we assume that the conditional distribution of $\bar{\beta}_{mj}(i)$ in (56) is approximately gaussian. Then, simplifying the expression in $\{\cdot\}$ of (56) gives the equation for $\tilde{\pi}_{mj}(i)$ as

$$\tilde{\pi}_{mj}(i) \cong \left[1 + \frac{1 - \tilde{\pi}_{mj}(i - 1)}{\tilde{\pi}_{mj}(i - 1)} \cdot \exp\left\{\frac{(-1)^y}{2}\right.\right.$$
$$\left.\left.\cdot[\bar{\beta}_{m1}(i) - \bar{\beta}_{m2}(i)]'\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1}[\bar{\beta}_{m1}(i) - \bar{\beta}_{m2}(i)]\right\}\right]^{-1}.$$
$$(57)$$

Now $\hat{\beta}_j$, defined in (A.4) is obtained similar to (51) as

$$E[\beta_l(k + 1)|Z^k, \beta^k, U^k, \theta = j] \cong$$

$$\phi_{lj}|\sin[\hat{\gamma}_l(k + 1|k) - \hat{\psi}_l(k + 1|k)]|$$

$$l = 1, 2 \qquad (58)$$

where

$$\hat{\gamma}_l(k + 1|k) \cong \tan^{-1}\frac{\dot{y}_l(k + 1|k) - \dot{y}_l(k + 1|k)}{\dot{x}_l(k + 1|k) - \dot{x}_l(k + 1|k)} \qquad (59)$$

and

$$\hat{\psi}_l(k + 1|k) \cong \tan^{-1}\frac{\dot{y}_l(k + 1|k)}{\dot{x}_l(k + 1|k)} \qquad (60)$$

with

$$\hat{x}_l(k + 1|k) = E[x_l(k + 1|Z^k, \beta^k, U^k] \qquad (61)$$

and

$$\hat{x}_l(k + 1|k) = E[x_l(k + 1)|Z^k, \beta^k, U^k] \qquad (62)$$

being obtained using Kalman filters.

Finally, S_j, the covariance matrix associated with (58) may be taken as

$$S_j \cong \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \qquad j = 1, 2. \qquad (63)$$

A final remark on the extension of the present work: In the analysis presented of the target/decoy interception problem, it was assumed that the target and the decoy will follow the same state models throughout, i.e. the models do not switch to other state models. After this algorithm of the target/decoy interception problem has been activated, if any of the vehicles do switch to a different state model, that switch must be detected and the corresponding filter should be reinitialized. Notice that the analysis presented in this work remains valid for the switched model as long as the states propagate according to an equation similar to (1), for example — a switch from a nearly constant speed (non-maneuvering) model of (41) to a nearly constant acceleration (maneuvering) model with the state vector

$$x = [x, \dot{x}, y, \dot{y}, \ddot{x}, \ddot{y}]'. \qquad (64)$$

A simple maneuver detection scheme for tracking a maneuvering target, i.e. a scheme to detect the switching of models, is given by Bar-Shalom and Birmiwal (1982). It was observed there that suitable state models at all times will result in the best tracking performance. Using such a scheme to detect the switching of models and then reinitializing the switched-state model, the present work is easily extended to the case of the target/decoy changing models.

6. SIMULATION RESULTS

As an evaluation of this algorithm, the above example was simulated. Two sets of feature lengths were chosen: one for the target and decoy being nearly 'identical' and the other corresponding to more separated features. For each set of features, two pairs of distinct trajectories for the two vehicles were considered. Initial values of these trajectories were

Trajectories 1

$$x_1(0) = \begin{bmatrix} 8000\,m \\ 100\,m/s \\ 500\,m \\ 0\,m/s \end{bmatrix}, x_2(0) = \begin{bmatrix} 8000\,m \\ 100\,m/s \\ -500\,m \\ 0\,m/s \end{bmatrix} \qquad (65)$$

Trajectories 2

$$x_1(0) = \begin{bmatrix} 7500\,m \\ 0\,m/s \\ 0\,m \\ 100\,m/s \end{bmatrix}, x_2(0) = \begin{bmatrix} 8500\,m \\ 0\,m/s \\ 0\,m \\ 100\,m/s \end{bmatrix}. \qquad (66)$$

The sampling time interval T was taken to be 3 sec. The process noise covariance matrix associated with the interceptor was taken to be zero while for the two vehicles, it was

$$Q_1 = Q_2 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}(m/s)^2. \qquad (67)$$

The interceptor's state measurement noise covariance, R_1 was taken to be zero (the interceptor knows its state with relatively more certainty and no Kalman filter for the interceptor) while for the two vehicles, it was

$$R_1 = R_2 = \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix}(m^2). \qquad (68)$$

The feature measurement noise variance was taken to be $\sigma_1^2 = \sigma_2^2 = 4\,m^2$. Since no information about target/decoy was available at time 0, $\pi(0) = 0.5$.

The cost matrix $R(i)$ associated with the interceptor control was taken to be the same for all i

$$R(i) = \begin{bmatrix} 10.0 & 0 \\ 0 & 10.0 \end{bmatrix}(s^2/m)^2 \quad \forall i. \qquad (69)$$

The cost matrix $\mathbf{Q}$ associated with the relative position of the terminal target decoy and interceptor state was taken to be

$$\mathbf{Q} = \begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \end{bmatrix} (m)^{-2}. \qquad (70)$$

The interceptor control bound $u_j^{max}(i)$ was taken to be $25\,m\,s^2$ for $j = 1,2$ and for all i. The discrete controls $u(k)$ were chosen over a grid of points (controls) $5\,m\,s^2$ apart in both the directions and whose effective direction of acceleration was within $90°$ of the direction of motion of the target/decoy. The speed limit of the interceptor. τ_1^{max} was taken to be $250\,m\,s$. The threshold π^*. which is used to decide about the identities of the tracks. was taken to be 0.499. i.e.. the decision about the tracks was made when $\pi(k)$ was greater than 0.999 or smaller than 0.001. After the decision about 'which track is from the target' is made. the bounded optimal control obtained from the solution of the LQG problem for the estimated time-to-go was applied to the interceptor until the determined target was intercepted.

Tracks of both vehicles were initialized using the two-point differencing of the measurements method. as in Bar-Shalom and Birmiwal (1982). Initial values of the interceptor state components were taken to be zero.

The two sets of feature values considered were $\phi_1 = 22\,m$. $\phi_2 = 20\,m$ and $\phi_1 = 28\,m$. $\phi_2 = 20\,m$. Observe that the features of the first set differ effectively by less than one standard deviation of the feature measurement noise.

For each of the four cases. a Monte Carlo simulation of ten runs was performed. It was observed that the interceptor intercepted the true target correctly in all the runs. Figure 2 shows the typical motions of the target. decoy and the interceptor. starting at time zero until the interception took place. for the very close features set and the trajectories one. Figures 3 5 show these
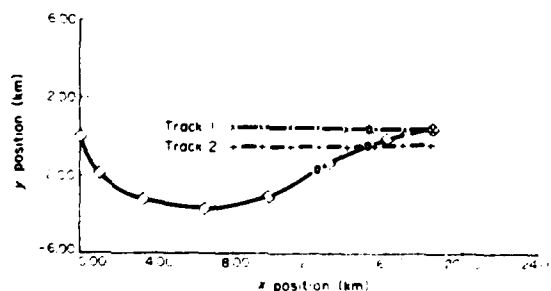
FIG. 2. Typical motions of the target. decoy and the interceptor for the case of $\theta = 1$. $\phi_1 = 22\,m$. $\phi_2 = 20\,m$ and trajectories 1. Here. the distance between two consecutive similar symbols is five sampling intervals (15 sec) and * represents the locations of the two vehicles and the interceptor when the decision about the tracks is made. Legend: × × target: + + decoy: ◇ interceptor.
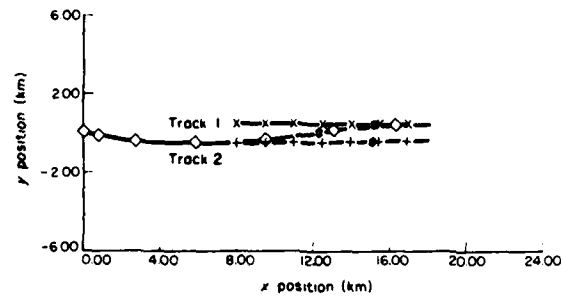
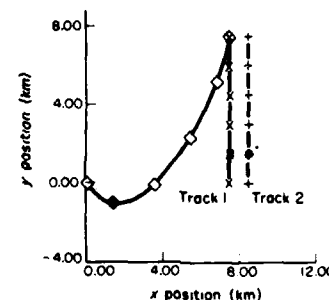FIG. 3. Typical motions for the case of $\theta = 1$. $\phi_1 = 28\,m$. $\phi_2 = 20\,m$ and trajectories 1.

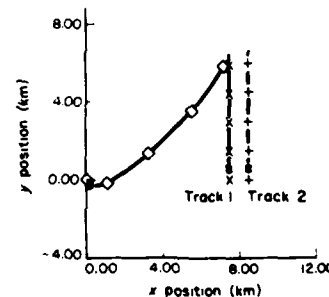FIG. 4. Typical motions for the case of $\theta = 1$. $\phi_1 = 22\,m$. $\phi_2 = 20\,m$ and trajectories 2.

FIG. 5. Typical motions for the case of $\theta = 1$. $\phi_1 = 28\,m$. $\phi_2 = 20\,m$ and trajectories 2.

motions for the other three cases. For the same set of random numbers and corresponding to each of the above 40 runs, another set of runs was performed with target and decoy tracks interchanged. Again, the true target was identified correctly and intercepted in all these runs. Figures 6–9 are the plots corresponding to Figs 2–5 respectively with θ changed (target and decoy switched).

From these figures, we observe that the interceptor takes longer time in deciding about the tracks when the interceptor is on the endfire than on the broadside. This is because the feature measurement noise is more dominating in the former case. When the target and decoy are more different, it takes less time to decide about the tracks, which is intuitively obvious. When the target and decoy are nearly identicial, the interceptor does not follow them directly. Instead, it takes a course so that at the

time of interception. the last nominal *a posteriori* probability $\bar{\pi}(N-1)$ is close to its extreme value (here we have the dual effect). To achieve this goal in minimum time. the interceptor tries to be on the broadside of the target/decoy. In case the target and decoy are easily discriminable. the interceptor follows the vehicles directly because it anticipates that the future learning will guide it correctly to the true target.
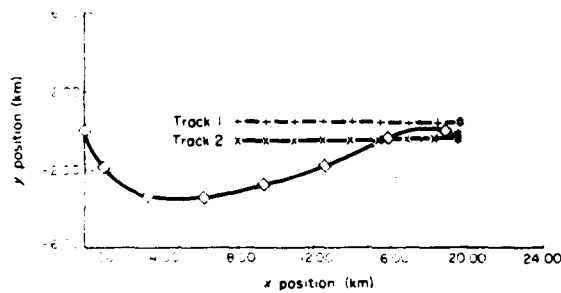
The relative importance of the terminal state cost over the interceptor control cost was seen by changing all diagonal components of **Q** to 0.05 m^{-2}. For this **Q** and the rest of the parameters unchanged, a Monte Carlo of ten runs was obtained for each set of the trajectories and features corresponding to Figs 2. 4, 6 and 8. It was observed that the true target was intercepted correctly in all these runs. Figures 10–13 are the respective plots giving the typical motion of
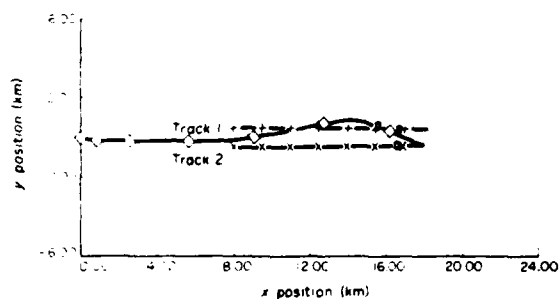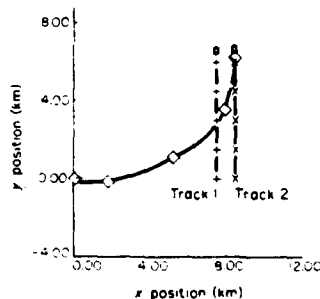

FIG. 6. Typical motions for the case of $\theta = 2$. $\phi_1 = 20$m. $\phi_2 = 22$m and trajectories 1.

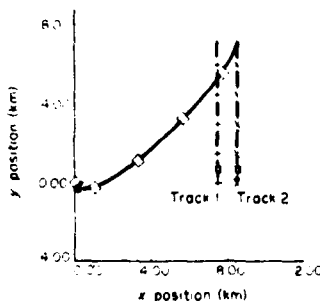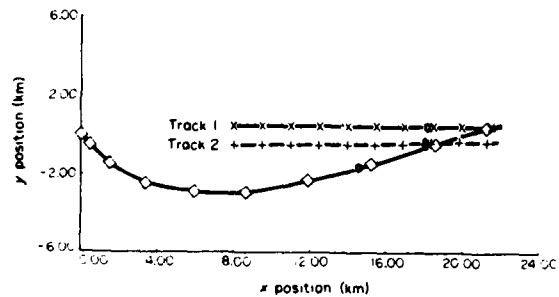FIG. 10. Typical motions for the case of $\theta = 1$, $\phi_1 = 22$m. $\phi_2 = 20$m, trajectories 1 and reduced **Q**.

FIG. 7. Typical motions for the case of $\theta = 2$. $\phi_1 = 20$m. $\phi_2 = 28$m and trajectories 1.

FIG. 11. Typical motions for the case of $\theta = 1$. $\phi_1 = 22$m. $\phi_2 = 20$m. trajectories 2 and reduced **Q**.

FIG. 8. Typical motions for the case of $\theta = 2$. $\phi_1 = 20$m. $\phi_2 = 22$m and trajectories 2.
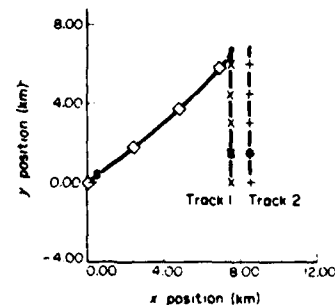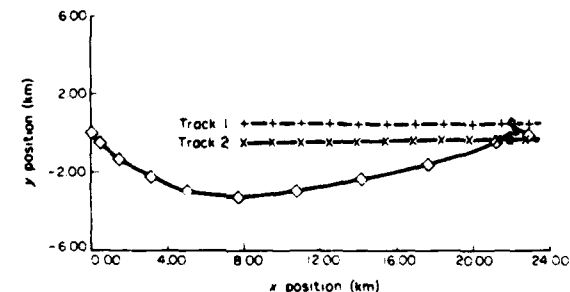

FIG. 12. Typical motions for the case of $\theta = 2$. $\phi_1 = 20$m. $\phi_2 = 22$m. trajectories 1 and reduced **Q**.

FIG. 9. Typical motions for the case of $\theta = 2$. $\phi_1 = 20$m. $\phi_2 = 28$m and trajectories 2.
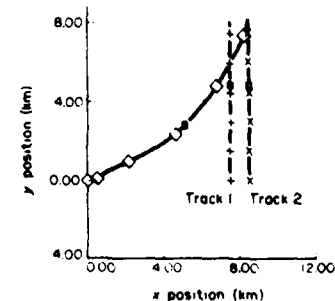

FIG. 13. Typical motions for the case of $\theta = 2$m, $\phi_1 = 20$m, $\phi_2 = 22$m, trajectories 2 and reduced **Q**.

the target, decoy and the interceptor. By comparing these two sets of figures, it is clear that the interceptor tried to be more on the broadside in the former case (and hence took less time to decide about the track ID) so that the last nominal posterior probability $\bar{\pi}(N-1)$ was closer to its extreme value. This is because the terminal state cost was relatively more dominating over the interceptor control cost in the former case than in the latter case, i.e. the controller was more willing to expend the additional fuel to take the more energetic trajectory in the former case and this resulted in faster convergence.

The algorithm was run in Fortran IV on IBM-3081D. The number of statements in the code were around 1300, but the code included overhead (trajectory, noise generation) and hence was not efficient. The memory requirement was approximately 250K and the average CPU time for each run was approximately 1 min.

7. CONCLUSIONS

An adaptive dual-control guidance algorithm for intercepting a moving target has been developed for the situation where the target is using a defensive decoy in a stochastic environment. At each time step, the interceptor chooses its bounded control and hence its trajectory such that it can differentiate between the true target of interest and the decoy with the aid of the expected future state observations and the feature measurements, approaching at the same time towards the target decoy. To reduce the computational load, an approximate solution of the stochastic dynamic programming equation is obtained by performing the preposterior analysis. The algorithm developed is especially useful if the cost associated with the terminal miss distance between the true target and the interceptor is relatively high compared to the interceptor control cost. The case of the target decoy changing their state models is also considered. The simulation studies have shown the effectiveness of the scheme.

REFERENCES

Bar-Shalom, Y. and K. Birmiwal (1982). Variable dimension filter for maneuvering target tracking. *I.E.E.E. Trans. Aerospace and Electronic Systems*, **18**, 621.

Bar-Shalom, Y. and F. Tse (1974). Dual effect, certainty equivalence, and separation in stochastic control. *I.E.E.E. Trans. Auto Cont.*, **AC-19**, 494.

Bar-Shalom, Y. and E. Tse (1976). Concepts and methods in stochastic control. In C. T. Leondes (Ed.), *Control and Dynamic Systems Advances in Theory and Applications*, Vol. 12. Academic Press, New York.

Bellman, R. (1961). *Adaptive Control Processes A Guided Tour*. Princeton University Press, Princeton, New Jersey.

Bryson, A. E. and Y. C. Ho (1969). *Applied Optimal Control*. Ch. 5. Ginn and Company, Waltham, Massachusetts.

Casler, Jr., R. J. (1978). Dual control guidance for homing interceptors with angle-only measurements. *J.A.J. Guidance and Control*, **1**, 63.

Chang, C. B. (1979). Ballistic trajectory estimation with angle-only measurement. *Proc. 18th IEEE Conf. on Decision and Control*, **2**, 632. Florida.

Deshpande, J. G., T. N. Upadhyay and D. G. Lainiotis (1973). Adaptive control of linear stochastic systems. *Automatica*, **9**, 107.

Dowdle, J. R., M. Athans, S. W. Gully and A. S. Willsky (1982). An optimal control and estimation algorithm for missile endgame guidance. *Proc. 21st IEEE Conf. on Decision and Control*, **3**, 1128. Florida.

Feldbaum, A. A. (1965). *Optimal Control Systems*, Academic Press, New York.

Lee, G. K. F. (1982). Investigation of time-to-go algorithms for air-to-air missiles. *Proc. 1st American Cont. Conf.*, **3**, 988. Virginia.

Murtaugh, S. A. and H. E. Criel (1966). Fundamentals of proportional navigation. *IEEE Spectrum*, December, 1975.

Newell, H. E., Jr. (1945). Guided missile kinematics. *Naval Res. Lab.*, Washington, D.C., Report R-2538.

Pastrick, H. L., S. M. Seltzer and M. E. Warren (1981). Guidance laws for short-range tactical missiles. *AIAA J. Guidance and Control*, **4**.

Raiffa, H. and R. Schlaifer (1972). *Applied Statistical Decision Theory*. M.I.T. Press, Cambridge, Massachusetts.

Riggs, T. (1979). An overview-optimal control and estimation for tactical missiles. *Nat. Aero. and Electr. Conf.*, p. 752.

Saridis, G. N. (1977). *Self-Organizing Control of Stochastic Systems*. Dekker, New York.

Speyer, J. L. and D. G. Hull (1982). Estimation enhancement by trajectory modulation for homing missiles. *Proc. 1st American Control Conf.*, **3**, 978. Virginia.

Tse, E. and Y. Bar-Shalom (1973). An actively adaptive control for discrete-time systems with random parameters. *IEEE Trans. Auto. Cont.*, **AC-18**, 109.

Tse, E. and Y. Bar-Shalom (1975). Adaptive dual control for stochastic nonlinear systems with free end-time. *IEEE Trans. Auto. Cont.*, **AC-20**, 670.

Wenk, C. J. and Y. Bar-Shalom (1980). A multiple model adaptive dual control algorithm for stochastic systems with unknown parameters. *IEEE Trans. Auto. Cont.*, **AC-25**, 703.

Wenk, C. J. (1981). Decision strategies with dual effect for inseparable stochastic control problems with continuous and discrete uncertainty. Ph.D. Thesis, University of Connecticut, Dept. of Electrical Engineering and Computer Science.

APPENDIX A: DERIVATION OF THE APPROXIMATE PREPOSTERIOR DENSITY

The locations $\pi_1(k+1)$ and $\pi_2(k+1)$ of the preposterior density (23) are obtained by matching the first two moments of (23) to the true preposterior moments of $\pi(k+1)$, viz. $E[\pi(k+1)|Z^k, \beta^k, U^k]$ and $E[\pi^2(k+1)|Z^k, \beta^k, U^k]$. From the Fundamental Theorem on Expectation and (19), we have

$$E[\pi(k+1)|Z^k, \beta^k, U^k] = \pi(k). \qquad (A.1)$$

Using the total probability theorem, the true second moment of $\pi(k+1)$ can be rewritten as

$$\bar{\pi}^2(k+1) \triangleq E[\pi^2(k+1)|Z^k, \beta^k, U^k] = \pi(k)$$
$$\cdot E[\pi^2(k+1)|Z^k, \beta^k, U^k, \theta = 1] + [1 - \pi(k)]$$
$$\cdot E[\pi^2(k+1)|Z^k, \beta^k, U^k, \theta = 2]. \qquad (A.2)$$

Now consider $E[\pi^2(k+1)|Z^k, \beta^k, U^k, \theta = j]$. In view of (19), and ignoring the variation of $\pi(k+1)$ with respect to $z(k+1)$, then $\pi(k+1)$ is an explicit function of $\beta(k+1)$

$$\pi(k+1) = \pi[\beta(k+1)] \triangleq \pi[\beta]. \qquad (A.3)$$

Expanding $\pi(k+1)$ to second order about

$$\beta_j \triangleq E[\beta(k+1)|Z^k, \beta^k, U^k, \theta = j] \qquad (A.4)$$

gives

$$\pi(k+1) \cong \pi_j + [\nabla\pi_j]'[\beta - \beta_j] + \tfrac{1}{2}[\beta - \beta_j]'[\nabla^2\pi_j][\beta - \beta_j]$$
$$(A.5)$$

where $\hat{\pi}_j, \nabla\hat{\pi}_j, \nabla^2\hat{\pi}_j$ are the function, gradient and Hessian of (A.3) respectively, evaluated at $\hat{\beta}_j$. Then, since $\beta - \hat{\beta}_j$ is approximately Gaussian, equation (A.5) gives

$$E[\pi^2(k+1)|Z^k,\beta^k,U^k,\theta=j] \cong [\hat{\pi}_j]^2 + \hat{\pi}_j\,\mathrm{tr}\,[(\nabla^2\hat{\pi}_j)S_j]$$

$$+ (\nabla\hat{\pi}_j)S_j(\nabla\hat{\pi}_j) + \tfrac{1}{2}\mathrm{tr}\,\{[(\nabla^2\hat{\pi}_j)S_j]^2\} + \tfrac{1}{4}\{\mathrm{tr}\,[(\nabla^2\hat{\pi}_j)S_j]\}^2 \quad (A.6)$$

where tr denotes the trace operator and $S_j = S(k+1|k,\theta=j)$ is the residual covariance associated with $\hat{\beta}_j$,

$$S(k+1|k,\theta=j) = \mathrm{cov}\,[\beta(k+1)|Z^k,\beta^k,U^k,\theta=j]. \quad (A.7)$$

This completes the computation of the second moment (A.2). Equating (A.1) and (A.2) to the respective moments of (23) gives

$$\pi(k) = \pi(k)\cdot\pi_1(k+1) + [1-\pi(k)]\cdot\pi_2(k+1) \quad (A.8)$$

$$\pi^2(k+1) = \pi(k)\cdot\pi_1^2(k+1) + [1-\pi(k)]\cdot\pi_2^2(k+1). \quad (A.9)$$

Using (24), (A.8) and (A.9) yields the desired two-point preposterior density locations

$$\pi_1(k+1) = \pi(k) + \left\{\frac{1-\pi(k)}{\pi(k)}\cdot[\pi^2(k+1)-\pi^2(k)]\right\}^{1/2} \quad (A.10)$$

$$\pi_2(k+1) = \frac{\pi(k)}{1-\pi(k)}\cdot[1-\pi_1(k+1)]. \quad (A.11)$$

APPENDIX B: THE ALGORITHM

Step 1
　Initialize $\pi(0)$ and $\hat{x}_1(0|0),\hat{x}_2(0|0),\hat{x}_1(0|0)$ along with their respective covariances. Obtain the predicted states $\hat{x}_1(1|0),\hat{x}_2(1|0)$. Define $k = 0$.

Step 2
　Is it desired to obtain the optimal control for target identification and interception (true only when $|\pi(k)-0.5| \le \pi^*$)? If yes, go to Step 3, otherwise terminate.

Step 3
　Choose a feasible control $u(k)$, i.e. a control that satisfies the constraints $|u_n(k)| \le u_n^{max}(k)$ and $\tau_1(k+1) \le \tau^{max}$. Obtain $\hat{x}_1(k+1|k)$.

$$\hat{x}_1(k+1|k) = A_1\hat{x}_1(k|k) + Bu(k). \quad (B.1)$$

Step 4
　Compute $\hat{\beta}_j = E[\beta(k+1)|Z^k,\beta^k,U^k,\theta=j]$ and the associated covariance S_j. For the example considered

$$\hat{\beta}_j \cong \begin{bmatrix} \phi_1|\sin\,[\zeta_1(k+1|k)-\tilde{\omega}_1(k+1|k)]| \\ \phi_2|\sin\,[\zeta_2(k+1|k)-\tilde{\omega}_2(k+1|k)]| \end{bmatrix} \quad (B.2)$$

where $\zeta_l(k+1|k),\tilde{\omega}_l(k+1|k); l = 1,2$ are given by the equations (59) and (60) respectively along with equations (61) and (62). The expression for S_j is given by (63).

Step 5
　Compute $\pi_1(k+1)$ and $\pi_2(k+1)$.

$$\pi_1(k+1) = \pi(k) + \left\{\frac{1-\pi(k)}{\pi(k)}\cdot[\pi^2(k+1)-\pi^2(k)]\right\}^{1/2} \quad (B.3)$$

$$\pi_2(k+1) = \frac{\pi(k)}{1-\pi(k)}\cdot[1-\pi_1(k+1)] \quad (B.4)$$

where $\pi^2(k+1) = E[\pi^2(k+1)|Z^k,\beta^k,U^k]$ is given by equation (A.2) which in turn is given by equation (A.6). This involves $\hat{\pi}_j$, $\nabla\hat{\pi}_j$ and $\nabla^2\hat{\pi}_j$, which, for the example considered, are given by

$$\hat{\pi}_j = \left[1 + \frac{1-\pi(k)}{\pi(k)}\cdot\exp\left\{-\frac{1}{2}\left[(\hat{\beta}_j-\hat{\beta}_2)\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}^{-1}(\hat{\beta}_j-\hat{\beta}_2)\right.\right.\right.$$
$$\left.\left.\left. - (\hat{\beta}_j-\hat{\beta}_1)\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}^{-1}(\hat{\beta}_j-\hat{\beta}_1)\right]\right\}\right]^{-1} \quad (B.5)$$

$$\nabla\hat{\pi}_j = \hat{\pi}_j\cdot(1-\hat{\pi}_j)\cdot\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}^{-1}\cdot(\hat{\beta}_j-\hat{\beta}_2) \quad (B.6)$$

$$\nabla^2\hat{\pi}_j = \hat{\pi}_j\cdot(1-\hat{\pi}_j)\cdot(1-2\hat{\pi}_j)\cdot\left\{\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}^{-1}\right.$$
$$\left.[\hat{\beta}_j-\hat{\beta}_2][\hat{\beta}_j-\hat{\beta}_2]\begin{bmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{bmatrix}^{-1}\right\}. \quad (B.7)$$

Step 6
　Compute the cost $J(k)$ for the control $u(k)$ as follows:

$$J(k) = u'(k)R(k)u(k) + E[J^*(k+1)|Z^k,\beta^k,U^k] \quad (B.8)$$

where

$$E[J^*(k+1)|Z^k,\beta^k,U^k] \cong \pi(k)\pi_1(k+1)$$
$$\hat{J}_{11}(k+1) + \pi(k)\cdot[1-\pi_1(k+1)]\hat{J}_{12}(k+1)$$
$$+ [1-\pi(k)]\cdot\pi_2(k+1)\hat{J}_{21}(k+1)$$
$$+ [1-\pi(k)]\cdot[1-\pi_2(k+1)]\cdot\hat{J}_{22}(k+1). \quad (B.9)$$

Now we compute $\hat{J}_{ml}(k+1)$ for the sequence mj, $m = 1,2$; $l = 1,2$.

(i) Estimate $\hat{N}(k)$. For the example considered

$$\hat{N}(k) = \min\{\hat{N}_1(k),\hat{N}_2(k)\}$$

where

$$\hat{N}_l(k) = k + \left[\frac{t_l^*}{T}\right] + 1 \quad l = 1,2 \quad (B.10)$$

where $[x]$ is the greatest integer less than or equal to x, T is the sampling interval and t_l^* is the positive root of the equation.

$$[(\tau_l^{max})^2 - \{(\dot{x}_l(k|k))^2 + (\dot{y}_l(k|k))^2\}]\cdot t^2$$
$$- 2[\hat{x}_l(k|k)\cdot(\dot{x}_l(k|k) - \dot{x}_l(k|k)) + \hat{y}_l(k|k)\cdot(\dot{y}_l(k|k)$$
$$- \dot{y}_l(k|k))]t - [(\hat{x}_l(k|k) - \hat{x}_l(k|k))^2$$
$$+ (\hat{y}_l(k|k) - \hat{y}_l(k|k))^2] = 0. \quad (B.11)$$

(ii) Define

$$\hat{N}_{ml}(k+1) = \hat{N}(k) \quad (B.12)$$

$$\hat{\pi}_{ml}(k+1) = \pi_m(k+1) \quad (B.13)$$

$$\hat{x}_{lm_l}(k+1) = A_l\hat{x}_l(k|k) \quad l = 1,2 \quad (B.14)$$

$$\hat{x}_{lm_j}(k+1) = A_l\hat{x}_l(k|k) + Bu(k). \quad (B.15)$$

The solution of the LQG problem for the example considered and for the nominal terminal time $\hat{N}_{m_j}(k+1)$ is

$$u_{lm_j}(k+1) = -\{[B'K_{m_j}(k+2)B + R(k+1)]^{-1}$$
$$B'K_{m_j}(k+2)A\}$$
$$\cdot[\hat{x}_{lm_j}(k+1) - \hat{x}_{lm_j}(k+1)] \quad l = 1,2 \quad (B.16)$$

where $K_{m_j}(k+2)$ is given by the backward equation (Ricatti equation)

$$K_{m_j}(k+1) = A'\{K_{m_j}(k+2) - K_{m_j}(k+2)B[B'K_{m_j}(k+2)B$$
$$\cdot R(k+1)]^{-1}B'K_{m_j}(k+2)\}A \quad (B.17)$$

$$K_{m_j}(\hat{N}_{m_j}(k+1)) = Q^a \quad (B.18)$$

and Q^a is the 4×4 matrix whose elements are all zero except the elements (1,1),(1,3),(3,1) and (3,3) which are the elements (1,1),

(1. 2), (2. 1) and (2. 2) respectively of the $\mathbf{Q}$ matrix. For the general case, the optimal control $\bar{\mathbf{u}}_{1m_j}(k+1)$ is a function of A_l, $l = 1, 2, A_l$ and the different state components of the three state vectors. The solution can be obtained for any specific problem by proper augmentation of the state vector and the transition matrices.

If the magnitude of the nth component of this control exceeds the bound $u_n^{max}(k+1)$, then change this component to $\pm u_n^{max}(k+1)$, i.e. sign (direction) unchanged. Obtain the DUL control

$$\bar{\mathbf{u}}_{m_j}(k+1) = \tilde{\pi}_{m_j}(k+1)\bar{\mathbf{u}}_{1m_j}(k+1) + [1 - \tilde{\pi}_{m_j}(k+1)]$$
$$\bar{\mathbf{u}}_{2m_j}(k+1). \qquad (B.19)$$

Truncate this control to $t \cdot \bar{\mathbf{u}}_{m_j}(k+1)$ $(0 < t \le 1)$, if necessary, so that the nominal speed of the interceptor at time $k+2$ is equal to τ_1^{max}.

(iii) For time $i \ge k+2$, define recursively

$$\bar{\mathbf{x}}_{1m_j}(i) = A_l\bar{\mathbf{x}}_{1m_j}(i-1) \qquad l = 1, 2 \qquad (B.20)$$

$$\bar{\mathbf{x}}_{m_j}(i) = A_l\bar{\mathbf{x}}_{m_j}(i-1) + B\bar{\mathbf{u}}_{m_j}(i-1) \qquad (B.21)$$

$$\beta_{1m_j}(i) = f[\phi_{l_j}, \bar{\mathbf{x}}_{1m_j}(i), \bar{\mathbf{x}}_{1m_j}(i)] \qquad l = 1, 2 \qquad (B.22)$$

$$\tilde{\pi}_{m_j}(i) = \left[1 + \frac{1 - \tilde{\pi}_{m_j}(i-1)}{\tilde{\pi}_{m_j}(i-1)} \right.$$
$$\left. \frac{p[\mathbf{z}_{m_j}(i), \beta_{m_j}(i)|Z_{m_j}^{-1}, \beta_{m_j}^{-1}, U_{m_j}^{-1}, \theta = 2]}{p[\mathbf{z}_{m_j}(i), \beta_{m_j}(i)|Z_{m_j}^{-1}, \beta_{m_j}^{-1}, U_{m_j}^{-1}, \theta = 1]} \right]^{-1} \qquad (B.23)$$

Now, estimate $\bar{N}_{m_j}(i)$ as was done for time k. If $\bar{N}_{m_j}(i) = i + 1$, then go to (iv). Otherwise continue. Obtain the DUL control

$$\bar{\mathbf{u}}_{m_j}(i) = \tilde{\pi}_{m_j}(i) \cdot \bar{\mathbf{u}}_{1m_j}(i) + [1 - \tilde{\pi}_{m_j}(i)] \cdot \bar{\mathbf{u}}_{2m_j}(i) \qquad (B.24)$$

where $\bar{\mathbf{u}}_{1m_j}(i)$ is obtained as for time $k+1$ and $\bar{\mathbf{u}}_{1m_j}(i)$ and $\bar{\mathbf{u}}_{m_j}(i)$ are adjusted, if necessary, as was done for time $k+1$. Go back to (iii).

For the example considered,

$$\beta_{1m_j}(i) = \phi_{l_j}|\sin[\gamma_{1m_j}(i) - \psi_{1m_j}(i)]| \qquad (B.25)$$

where

$$\gamma_{1m_j}(i) = \tan^{-1}\frac{\bar{y}_{1m_j}(i) - \bar{y}_{1m_j}(i)}{\bar{x}_{1m_j}(i) - \bar{x}_{1m_j}(i)} \qquad (B.26)$$

$$\psi_{1m_j}(i) = \tan^{-1}\frac{\bar{y}_{1m_j}(i)}{\bar{x}_{1m_j}(i)} \qquad (B.27)$$

and $\tilde{\pi}_{m_j}(i)$ is given by the equation (57).

(iv) $$\bar{J}_{m_j}(k+1) \cong \sum_{s=k+1}^{i-1} \bar{\mathbf{u}}_{m_j}'(s)\mathbf{R}(s)\bar{\mathbf{u}}_{m_j}(s)$$
$$+ [\tilde{\pi}_{m_j}(i) \cdot \mathbf{g}'[\bar{\mathbf{x}}_{1m_j}(i+1), \bar{\mathbf{x}}_{1m_j}(i+1)]\mathbf{Q}\mathbf{g}[\bar{\mathbf{x}}_{1m_j}(i+1),$$
$$\bar{\mathbf{x}}_{1m_j}(i+1)] + [1 - \tilde{\pi}_{m_j}(i)] \cdot \mathbf{g}'[\bar{\mathbf{x}}_{2m_j}(i+1),$$
$$\bar{\mathbf{x}}_{1m_j}(i+1)]\mathbf{Q} \cdot \mathbf{g}[\bar{\mathbf{x}}_{2m_j}(i+1), \bar{\mathbf{x}}_{1m_j}(i+1)]]. \qquad (B.28)$$

For the example considered,

$$\bar{J}_{m_j}(k+1) \cong \sum_{s=k+1}^{i-1} \bar{\mathbf{u}}_{m_j}'(s)\mathbf{R}(s)\bar{\mathbf{u}}_{m_j}(s)$$
$$+ \tilde{\pi}_{m_j}(i) \cdot \{[\bar{x}_{1m_j}(i+1) - \bar{x}_{1m_j}(i+1)]^2$$
$$+ [\bar{y}_{1m_j}(i+1) - \bar{y}_{1m_j}(i+1)]^2\}^{1/2}$$
$$+ [1 - \tilde{\pi}_{m_j}(i)]\{[\bar{x}_{2m_j}(i+1) - \bar{x}_{1m_j}(i+1)]^2$$
$$+ [\bar{y}_{1m_j}(i+1) - \bar{y}_{1m_j}(i+1)]^2\}^{1/2}. \qquad (B.29)$$

Step 7

Find the cost $J(k)$ for different discrete controls and find the control $\mathbf{u}^*(k)$ that minimizes $J(k)$. Apply $u^*(k)$ at time k.

Step 8

Obtain the observations $\mathbf{z}_l(k+1)$, $\beta_l(k+1)$, $l = 1, 2$. and $z_l(k+1)$. Update the estimates $\hat{\mathbf{x}}_l(k+1|k+1)$, $\hat{\mathbf{x}}_l(k+1|k+1)$, $\hat{\mathbf{x}}_l(k+2|k+1)$ using Kalman filters. Update the posterior probability $\pi(k+1)$ using equation (49) and for the example considered, using equation (50). Increment k by one and go to step 2.

DUAL ADAPTIVE CONTROL BASED UPON SENSITIVITY FUNCTIONS*

J. A. Molusis, P. Mookerjee, Y. Bar-Shalom

Dept. of Electrical Engineering & Computer Science
The University of Connecticut
Storrs, CT 06268

ABSTRACT

A new adaptive dual control solution is presented
for the control of a class of multi-variable input-
output systems. Both rapidly varying random parameters
and constant but unknown parameters are included. The
new controller modifies the cautious control design by
numerator and denominator correction terms. This con-
troller is shown to depend upon sensitivity functions
of the expected future cost. A scalar example is pre-
sented to provide insight into the properties of the
new dual controller. Monte-Carlo simulations are per-
formed which show improvement over the cautious con-
troller and the Linear Feedback Dual Controller of
[1] and [2].

1. INTRODUCTION

Multi-variable systems which are characterized by
uncertain parameters with large random variations are a
difficult challenge for most control design techniques.
The assumed randomness of the parameter variations
often precludes the use of gain scheduling (non adapt-
ive) control design. *Stochastic adaptive control*
theory provides a principal design approach for systems
of this type. Exact solution of the stochastic prob-
lem with unknown parameters requires solution of the
Stochastic Dynamic Programming equation and this is not
feasible for practical implementation. The solution is
known to have a dual effect [1,2] that can be used to
enhance the real-time identification of system paramet-
ers as well as provide good control.

Many suboptimal dual solutions have been suggested
[1,2,5-11]. The various approaches which have incor-
porated this dual property can be loosely divided into
two classes. In the first class [5-8], the optimal
control problem is reformulated to consist of a one-
step ahead criterion to be minimized, augmented by a
second term which penalizes the cost for poor identifi-
cation. This approach is attractive due to the analy-
tical tractability of the solution; however, the solu-
tion is based on a one-step criterion and does not
fully exploit the dual property of a multi-step solu-
tion. Padilla and Cruz [14] give a dual control solu-
tion for such a plant by minimizing the control object-
ive function subject to an upper bound in the total
estimation cost. Their objective function includes a
standard control objective function and also a second
constraint term which reflects the sensitivity of the
parameters to the state of the system. Thus the solu-
tion adjusts itself to exercise better estimation for
such sensitive parameters within the upper bound. The
second class [9-11] utilizes the stochastic dynamic
programming equation directly and performs lineariza-
tion of the future cost in order to obtain a solution.
Previous control solutions among this second class re-
quire a numerical search procedure which poses diffi-
culties for a practical solution for on-line control
for multivariable systems.

The linear feedback dual controller of [1,2] is

based upon a first order Taylor series expansion of the
expected future cost and is called the first order dual
(FOD). It offers some improvement over the non dual
cautious control based upon a one-step criterion. The
results are based upon a simulation model with constant
but unknown parameters. Although the dual control of-
fers some improvement over the cautious controller the
improvement is not significant for most practical ap-
plications where the system contains constant parameters
and the objective is to control in steady state opera-
tion. However, for random parameter variations, dual
control can sometimes offer significant improvement
over non-dual controllers [5,9]. The FOD of [1,2] is
attractive due to its simplicity (it is comparable to
the cautious control design in algorithm complexity and
does not require numerical search). The objective of
the present study is to evaluate the cautious control-
ler and the FOD for large random parameter variations
modeled as a random walk. Monte-Carlo simulations are
performed and conditions quantified under which the
dual controller offers significant improvement over a
non-dual cautious controller.

The FOD, although offering a reduction in the aver-
age cost, is found to be unacceptable in many cases.
This is attributed to the sensitivity of the expected
future cost whenever the system is characterized by
limited controllability. A second order expansion of
the linearization procedure of [1,2] is presented to
account for this sensitivity. This new second order
dual controller (SOD) inherently includes a robustness
property in that the controller accounts for sensiti-
vity of the expected future cost due to parameter esti-
mates and their uncertainty. Simulations are presented
which show the improvement of the SOD over the cautious
controller and the FOD. This SOD uses a Newton type
search procedure and is developed for multi-variable
systems. One of the main advantages of the SOD pre-
sented herein is that it modifies the cautious control-
ler with a numerator "probing" term and a denominator
correction term. Although the SOD is still considered
too complex for practical implementation, the structure
of the control solution is in a form which permits
practical design changes to the cautious controller to
include the dual properties.

Section 2 gives the problem formulation. The ap-
proximate dual controller for the multi-variable input-
output system is developed in Section 3. Section 4
analyzes this dual controller for a scalar example with
one unknown parameter. Section 5 concludes the paper.

2. PROBLEM FORMULATION

The multivariable system under investigation is

$$x(k+1) = c(k) + B(k) u(k) \qquad (2.1)$$

where $c(k)$ is an unknown vector and $B(k)$ is a matrix of
unknown parameters. The unknown elements of $c(k)$ and
$B(k)$ are denoted as $\theta(k)$ with covariance matrix $P(k)$.
These are represented by a discrete random model

$$\theta(k+1) = A\theta(k) + v(k) \qquad (2.2)$$

$$E(v(k))=0 \quad \text{and} \quad E(v(k)v'(j)) = V \delta_{kj} \qquad (2.3)$$

The measurement equation is

$$y(k) = x(k) + w(k) \qquad (2.4)$$

where

$$E(w(k)) = 0 \quad \text{and} \quad E(w(k)w'(j)) = W\,\delta_{kj}$$
$$E(w(k)v'(j)) = 0 \qquad (2.5)$$

and $x(k)$, $y(k)$ being n dimensional vectors. The control criterion to be minimized is the expected value of the cost from step 0 to N

$$J(0) = E\{C(0)\} = E\left\{ \sum_{k=1}^{N} x'(k)Qx(k) + u'(k-1)Ru(k-1) \right\} \qquad (2.6)$$

where $N = 2$ for the two step ahead criterion.

3. APPROXIMATE DUAL CONTROLLER FOR TWO STEP CRITERION

The minimization of (2.6) with respect to $u(0)$ and $u(1)$ subject to (2.1) - (2.5) is obtained from the Stochastic Dynamic Programming equation [12,13]

$$J^*(k) = \min_{u(k)} E\{C(k) + J^*(k+1) | Y^k\} \quad k=N-1,\ldots,1,0 \qquad (3.1)$$

where $J^*(k)$ is the "cost-to-go" from k to N and Y^k is the cumulated information at time k when the control $u(k)$ is to be determined. For $N = 1$, (3.1) is

$$J^*(0) = \min_{u(0)} E\{x'(1)Qx(1) + u'(0)Ru(0) + J^*(1) | Y^0\} \qquad (3.2)$$

where $J^*(1)$ is the optimal cost at the last step and is obtained by minimization of $J(N-1)$ for $N = 2$. Assuming diagonal $Q = \text{diag}(q_\ell)$ this results in [1,2]

$$J^*(1) = \hat{c}'(1)Q\hat{c}(1) + \sum_{\ell=1}^{n} q_\ell P_c^\ell(1) \qquad (3.3)$$

$$- [\hat{c}'(1)Q\hat{B}(1) + \sum_{\ell=1}^{n} q_\ell P_{cB}^\ell(1)] \, [\hat{B}'(1)Q\hat{B}(1) +$$

$$+ \sum_{\ell=1}^{n} q_\ell P_B^\ell(1) + R]^{-1} \cdot [\hat{B}'(1)Q\hat{c}(1) + \sum_{\ell=1}^{n} q_\ell P_{Bc}^\ell(1)]$$

and

$$u^*(1) = -[\hat{B}'(1)Q\hat{B}(1) + \sum_{\ell=1}^{n} q_\ell P_B^\ell(1) + R]^{-1}[\hat{B}'(1)Q\hat{c}(1)$$

$$+ \sum_{\ell=1}^{n} q_\ell P_{Bc}^\ell(1)] \qquad (3.4)$$

where

$$P^\ell(1) = \begin{bmatrix} P_c^\ell(1) & P_{cB}^\ell(1) \\ P_{Bc}^\ell(1) & P_B^\ell(1) \end{bmatrix} \qquad (3.5)$$

$P(1)$ is the expected value of $(\theta(1))^2$ for time step 2 given measurement $y(1)$ at time step 1. The index ℓ is used to represent the row number in (2.1) and $P^\ell(1)$ is the associated parameter covariance.

The parameter estimates $\hat{\theta}(1)$ and covariances $P(1)$ are obtained from the Kalman filter. Since W is diagonal one can decouple the estimation. Then

$$\hat{\theta}^\ell(1) = A\hat{\theta}^\ell(0) + AK_\ell(1)\,\nu_\ell(1) \qquad (3.6)$$

$$K_\ell(1) = P^\ell(0)H'(1)[H(1)P^\ell(0)H'(1) + W_\ell]^{-1} \qquad (3.7)$$

$$\tilde{P}^\ell(1) = P^\ell(0) - K_\ell(1)H(1)P^\ell(0) \qquad (3.8)$$

$$P^\ell(1) = A\tilde{P}^\ell(1)A' + V \qquad (3.9)$$

where

$$\nu_\ell(1) = y_\ell(1) - H(1)\hat{\theta}^\ell(0) \qquad (3.10)$$

$$H(1) = [1 \; u^T(0)] \qquad (3.11)$$

$$\theta^\ell(1) = [c_\ell(1) \; B_\ell(1)]^T, \quad \ell=1,2,\ldots n \text{ row of B} \qquad (3.12)$$

As discussed in [1] and [2] $J^*(1)$ is a nonlinear function of the parameter estimates $\hat{\theta}(1)$ and covariances $P(1)$ and thus a linearization was performed. In [1] a scalar formulation was presented and a first order linearization was performed about the nominal parameter estimate squared $(\bar{\theta}(0))^2$ and nominal covariance $\bar{P}(1)$. Also in [1,2] the vector case was presented and linearization to first order performed. To more accurately account for the dual effect a second order Taylor Series expansion is presented about $\bar{\theta}(0)$ and a first order expansion about the nominal covariance $\bar{P}(1)$. In addition (as will be presented subsequently) the covariance $P(1)$ will include a linearization to second order in $u(0)$. In [1,2], $P(1)$ was linearized to first order. It is believed that linearizations to second order are necessary to better account for the nonlinearity in $P(1)$ and $\theta(1)$ of (3.3) and in $u(0)$ of (3.7) and (3.8). In addition a nonlinear Newton algorithm is used in the second order approximation.

Linearization of (3.3) about the nominal $\bar{\theta}(1) = A\bar{\theta}(0)$ and $\bar{P}(1)$ using the nominal $\bar{u}(0)$ results in

$$J^*(1) = J^*[1, \hat{\theta}(0), \bar{P}(1)] + \frac{\partial J^*(1)}{\partial\hat{\theta}(1)} [\hat{\theta}(1) - A\hat{\theta}(0)]$$

$$+ \frac{1}{2}[\hat{\theta}(1) - A\hat{\theta}(0)]' \frac{\partial^2 J^*(1)}{\partial\hat{\theta}^2(1)} [\hat{\theta}(1) - A\hat{\theta}(0)]$$

$$+ \sum_{\ell=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)} [P_{i,j}^\ell(1) - \bar{P}_{i,j}^\ell(1)] \qquad (3.13)$$

where the superscript ℓ represents the covariance matrix associated with the ℓth row of parameters and $P_{i,j}(1)$ is the i-j th element of the covariance matrix $P(1)$, m being the number of unknown parameters.

Using (3.6) the expected value of (3.13) is

$$E[J^*(1) | Y^0] = J^*[1, \hat{\theta}(0), \bar{P}(1)]$$

$$+ \frac{1}{2} \text{tr}[\frac{\partial^2 J^*(1)}{\partial\hat{\theta}^2(1)} K(1) E\{\nu(1)\nu'(1) | Y^0\}K'(1)]$$

$$+ \sum_{\ell=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)} [P_{i,j}^\ell(1) - \bar{P}_{i,j}^\ell(1)] \qquad (3.14)$$

Using (3.7), (3.8) and the innovation covariance

$$E\{\nu_\ell(1)\,\nu_\ell'(1) | Y^0\} = H(1)P^\ell(0)H'(1) + W_\ell \qquad (3.15)$$

(3.14) can be written as

$$E[J^*(1) | Y^0] = J^*[1, \hat{\theta}(0), \bar{P}(1)]$$

$$+ \sum_{\ell=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \left\{ -\frac{1}{2} \frac{\partial}{\partial\hat{\theta}_i(1)} \frac{\partial J^*(1)}{\partial\hat{\theta}_j(1)} [P_{i,j}^\ell(1) - AP_{i,j}^\ell(0)A' - V_{i,j}^\ell] + \frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)} [P_{i,j}^\ell(1) - \bar{P}_{i,j}^\ell(1)] \right\} \qquad (3.16)$$

The expected future cost (3.16) is shown to be a function of the predicted covariance $P_{i,j}^\ell(1)$ with a multiplier given by the sensitivity $\frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)}$ and $\frac{\partial}{\partial\hat{\theta}_i(1)} \frac{\partial J^*(1)}{\partial\hat{\theta}_j(1)}$. Since the covariance $P_{i,j}^\ell(1)$ depends on the control $u(0)$ the control has the dual effect. It should be noted that the importance of the dual effect depends upon the sensitivity of the expected future cost with respect to both the covariance and parameter estimate.

The optimal control $u(0)$ can be computed by minimization of (3.2) using (3.16). Since $P_{i,j}^\ell(1)$ is nonlinear in $u(0)$ a numerical search procedure is required. This is accomplished using a second order linearization

2

in u(0).

Thus (3.8) is linearized to second order about the control $u^I(0)$, which is in the vicinity of the optimal control.

$$P_{i,j}^{\ell}(1) \approx \bar{P}_{i,j}^{\ell}(1) + \frac{\partial P_{i,j}^{\ell}(1)}{\partial u(0)}\Bigg|_{u^I(0)} [u(0)-u^I(0)]$$

$$+ \frac{1}{2}[u(0)-u^I(0)]' \frac{\partial^2 P_{i,j}^{\ell}(1)}{\partial u^2(0)}\Bigg|_{u^I(0)} [u(0)-u^I(0)] \qquad (3.17)$$

The expected future cost as given by (3.16) and (3.17) is quadratic in u(0) and thus a closed form solution $u^*(0)$ is obtained by minimization of (3.2).

The optimal dual control $u^*(0)$ can now be computed from (3.2) using (3.16) and (3.17). It is obtained by solving

$$\qquad (3.18)$$

$$\frac{\partial}{\partial u(0)} E\{x'(1)Q\,x(1) + u'(0)Ru(0) + J^*(1)|Y^0\} = 0$$

The optimal $u^*(0)$ is thus

$$u^*(0) = -[\hat{B}'(0)Q\hat{B}(0) + \sum_{\ell=1}^{n}(q_\ell P_B^\ell(0) + F_\ell) + R]^{-1}$$

$$[\hat{B}'(0)Q\hat{c}(0) + \sum_{\ell=1}^{n}(q_\ell P_{Bc}^\ell(0) + f_\ell)] \qquad (3.19)$$

where the matrix F_ℓ and the vector f_ℓ are

$$F_\ell = \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{2}\left(\frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)} - \frac{1}{2}\frac{\partial}{\partial\hat{\theta}_i^\ell(1)}\frac{\partial J^*(1)}{\partial\hat{\theta}_j^\ell(1)}\right)$$

$$\frac{\partial}{\partial u(0)}\frac{\partial P_{i,j}^\ell(1)}{\partial u(0)}\Bigg|_{u^I(0),\hat{\theta}(0),\bar{P}(1)} \qquad (3.20)$$

$$f_\ell = \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{2}\left(\frac{\partial J^*(1)}{\partial P_{i,j}^\ell(1)} - \frac{1}{2}\frac{\partial}{\partial\hat{\theta}_i^\ell(1)}\frac{\partial J^*(1)}{\partial\hat{\theta}_j^\ell(1)}\right)\left(\frac{\partial P_{i,j}^\ell(1)}{\partial u(0)} - \right.$$

$$\left. - \frac{\partial}{\partial u(0)}\frac{\partial P_{i,j}^\ell(1)}{\partial u(0)} u^I(0)\right)\Bigg|_{u^I(0),\hat{\theta}(0),\bar{P}(1)} \qquad (3.21)$$

Initially the nominal value of $\bar{u}(0)$ is computed from (3.19) with F_ℓ and f_ℓ equal to zero. Then a gradient search is performed until in the vicinity of the optimal $u^*(0)$. Then (3.19) - (3.21) are used until convergence is achieved. This iteration procedure is essentially Newton's method for minimization of a non-linear function. The gradient search is used because the stochastic cost in (3.2) being minimized is a high order nonlinear equation and the gradient procedure is used until $u^I(0)$ is in the vicinity of the minimum before switching to the Newton method. The nominal covariance $\bar{P}^\ell(1)$ is computed from((3.7) - (3.11)) with $u(0) = \bar{u}(0)$. The sensitivity (partials) in (3.20) and (3.21) of the cost $J^*(1)$ are computed from partial derivatives of the cost $J^*(1)$ (3.3) and $P^\ell(1)$ (3.7) - (3.9) evaluated at the nominal. The partials of the covariance are evaluated at $u^I(0)$ which is evaluated at the previous iteration I.

The approximate two-step ahead dual control of (3.19) - (3.21) can be interpreted as a modification to the cautious controller by the terms F_ℓ and f_ℓ. These terms depend upon the sensitivity of the future nominal cost $J^*(1)$ with respect to the parameters $\hat{\theta}_i^\ell(1)\hat{\theta}_j^\ell(1)$ for all i,j and their covariance $P_{i,j}^\ell(1)$ for each row ℓ of parameters. Whenever these sensitivities are large the terms F_ℓ and f_ℓ will be significant (that is the dual effect will be important). Thus the sensitivities take into account in the control solution the sensitivity of the nominal future cost due to parameter variation and uncertainty. The larger this sensitivity

the more important will be the dual effect.

The resulting dual controller (3.19) exhibits a robustness property with respect to parameter variations and uncertainty of the future cost by including a term which appears in the denominator of the dual controller. In addition, a probing term appears in the numerator.

4. SCALAR EXAMPLE WITH ONE UNKNOWN PARAMETER

To further understand the dual control solution a scalar example with one unknown parameter b is presented. The approximate dual control solution for this scalar case using Q = 1, R = 0, is given by (3.19) - (3.21) with $P_{i,j}(1)$ and $\hat{\theta}(0)$ being replaced by $P_b(1)$ and $\hat{b}(0)$ respectively.

The partials required in the control law are

$$\frac{\partial J^*(1)}{\partial P_b(1)}\Bigg|_{\hat{b}(0),\bar{P}_b(1)} = \frac{c^2\hat{b}^2(0)}{(\hat{b}^2(0)+\bar{P}_b(1))^2} \qquad (4.1)$$

$$\frac{\partial^2 J^*(1)}{\partial\hat{b}(1)\partial\hat{b}(1)}\Bigg|_{\hat{b}(0),\bar{P}_b(1)} = -2c^2\bar{P}_b(1)\left|\frac{\bar{P}_b(1)-3\hat{b}^2(0)}{(\hat{b}^2(0)+\bar{P}_b(1))^3}\right| \qquad (4.2)$$

$$\frac{\partial P_b(1)}{\partial u(0)}\Bigg|_{u^I(0)} = -\frac{2P_b^2(0)W\,u^I(0)a^2}{(P_b(0)u^{I^2}(0)+W)^2} \qquad (4.3)$$

$$\frac{\partial^2 P_b(1)}{\partial u(0)\partial u(0)}\Bigg|_{u^I(0)} = -2P_b^2(0)W\left|\frac{W-3P_b(0)u^{I^2}(0)}{(P_b(0)u^{I^2}(0)+W)^3}\right|a^2 \qquad (4.4)$$

where the nominal $\bar{u}(0)$ and $\bar{P}_b(1)$ are

$$\bar{u}(0) = -\frac{\hat{b}(0)c}{\hat{b}^2(0)+P_b(0)} \qquad (4.5)$$

$$\bar{P}_b(1) = \frac{a^2 P_b(0)W}{P_b(0)\bar{u}^2(0)+W} + V \qquad (4.6)$$

The parameter estimate $\hat{b}(0)$ and $P_b(0)$ are computed using data up to k = 0 (i.e. y(0)).

The expected future cost based upon the linearization of (3.16) is

$$E\{J^*(1)|Y^0\} = c^2 - c^2\frac{\hat{b}^2(0)}{\hat{b}^2(0)+\bar{P}_b(1)} - \frac{1}{2}\frac{\partial^2 J^*(1)}{\partial\hat{b}^2(1)}(P_b(1)$$

$$- a^2 P_b(0)-V) + \frac{\partial J^*(1)}{\partial P_b(1)}(P_b(1)-\bar{P}_b(1)) \qquad (4.7)$$

4.1 Evaluation of the Cautious Controller

The performance of the cautious controller can be evaluated using (3.2) with u(0) evaluated at the nominal

$$J(0) = [E\{x^2(1)|Y^0\} + E\{J^*(1)|Y^0\}]_{u(0)=\bar{u}(0)} \qquad (4.8)$$

The first term in (4.8) represents the expected cost at k = 1 and the second term in (4.8) represents the expected future cost at k = 2 using the cautious control at k = 1 (i.e. u(1)) and using the cautious control at k = 1 (i.e. u(0)=$\bar{u}$(0)). (4.8) is evaluated using data Y^0.

Using (4.1) - (4.7), (4.8) becomes,

$$J(0) = c^2 - \frac{\hat{b}^2(0)c^2}{\hat{b}^2(0)+P_b(0)} + c^2 - c^2\frac{\hat{b}^2(0)}{\hat{b}^2(0)+P_b(1)}$$

$$+ \frac{1}{2}\frac{\partial^2 J^*(1)}{\partial\hat{b}^2(1)}\cdot\frac{a^2 P_b^2(0)\bar{u}^2(0)}{P_b(0)\bar{u}^2(0)+W} \qquad (4.9)$$

3

The last term in (4.7) is zero since $P_b(1)$ evaluated at the nominal control (i.e. cautious control) equals $\bar{P}_b(1)$. The first two terms in (4.9) represent the average cost at step $k = 1$ and the last three terms represent the expected future cost at $k = 2$ using the cautious control.

A simple example can be used with (4.9) to demonstrate when the cautious control is expected to behave poorly.

Assume a scalar example with one unknown b parameter and let

$$\hat{b}(0) = .05 \ , \ P(0) = .5 \ , \ a = 1.0$$
$$V = .1 \ , \ W = .1 \ , \ c = 1 \tag{4.10}$$

The expected cost at $k = 1$ and $k = 2$ is computed from the nominal, $\bar{u}(0)$, $\bar{P}_b(1)$ and $\dfrac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)}$ which yields

$$\bar{u}(0) = -.1 \ , \ \bar{P}_b(1) = .575 \ , \ \frac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)} = -3.47 \tag{4.11}$$

and

$$J(0) \approx c^2 + c^2 \ , \ c = 1 \tag{4.12}$$

Thus the cautious control applied at $k = 0$ results in no reduction in the cost at $k = 1$ due to large uncertainty $P(1)$ and also no reduction in the future expected cost since $\bar{u}(0)$ is small and no improvement in parameter accuracy occurs at step $k = 1$.

4.2 Evaluation of the Dual Controller

The dual controller of (3.19) - (3.21), (4.1) - (4.6) can be evaluated by computing the average cost of (4.8) using the covariance

$$P_b(1) = \frac{a^2 P_b(0) W}{P_b(0) u^2(0) + W} + V \tag{4.13}$$

The expected future cost (4.7) reduces to

$$E\{J^*(1)|Y^o\}\Big|_{u^*(0)} = c^2 - c^2 \frac{\hat{b}^2(0)}{\hat{b}^2(0) + \bar{P}_b(1)}$$

$$+ \frac{1}{2} \frac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)} \frac{a^2 P_b^2(0) u^{*2}(0)}{P_b(0) u^{*2}(0) + W}$$

$$- \frac{\partial J^*(1)}{\partial P_b(1)} \left(\frac{a^2 P_b^2(0) u^{*2}(0)}{P_b(0) u^{*2}(0) + W} - \frac{a^2 P_b^2(0) \bar{u}^2(0)}{P_b(0) \bar{u}^2(0) + W} \right) \tag{4.14}$$

and the total expected cost at $k = 1$ and $k = 2$ using (4.8) is

$$J^*(0) = E\{x^2(1)|Y^o\}\Big|_{u^*(0)} + E\{J^*(1)|Y^o\}\Big|_{u^*(0)} \tag{4.15}$$

where

$$E\{x^2(1)|Y^o\}\Big|_{u^*(0)} = c^2 + 2\hat{b}(0) u^*(0) c +$$

$$+ (\hat{b}^2(0) + P_b(0)) u^{*2}(0) \tag{4.16}$$

Examination of (4.14) shows that the dual control can reduce the expected future cost over the cautious control since the last two expressions in (4.14) can be negative if $u^{*2}(0) > \bar{u}^2(0)$. Thus the dual property can have a desirable effect on the future cost.

The cost $J^*(0)$ is computed using the scalar example previously discussed for the cautious controller. A search procedure is used on (4.15) using (4.14) and (4.16) with the parameter values from (4.10), and $u^*(0)$ is iterated until in the vicinity of the minimum yielding

$$\frac{\partial J^*(1)}{\partial P_b(1)}\Big|_{\bar{u}(0)=-.1} = .0075 \ , \ \frac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)}\Big|_{\bar{u}(0)=-.1} = -3.47 \ ,$$

$$\frac{\partial P_b(1)}{\partial u(0)}\Big|_{u^I(0)=-.6} = .382 \ , \ \frac{\partial^2 P_b(1)}{\partial u^2(0)}\Big|_{u^I(0)=-.6} = +1.0 \ ,$$

$$F_\ell = .87 \ , \ f_\ell = .85 \tag{4.17}$$

The above sensitivities (4.17) were evaluated in the vicinity of the optimal $u^I(0) = -.6$ and $P_b(1) = .278$. The dual control $u^*(0)$ using $u^I(0) = -.6$, $c = 1$ is

$$u^*(0) = -\frac{\hat{b}(0)c + .85}{\hat{b}^2(0) + P_b(0) + .87} = -.62 \tag{4.18}$$

The corresponding future expected cost using (4.14) and (4.17) is

$$E\{J^*(1)|Y^o\}\Big|_{u^*(0)} \approx c^2 + \frac{1}{2} \frac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)} \frac{P_b^2(0) u^{*2}(0)}{P_b(0) u^{*2}(0) + W}$$

$$= .442 \ c^2 \ , \ c = 1 \tag{4.19}$$

The result of this example shows that the dual control of (4.18) reduces the expected future cost to 44% of the original c^2 with no control. The cautious control resulted in no reduction of the future cost. The terms responsible for the improvement with dual control are the second order sensitivities $\dfrac{\partial^2 P(1)}{\partial u^2(0)}$ and $\dfrac{\partial^2 J^*(1)}{\partial \hat{b}^2(1)}$.

The dual control of (4.18) differs from the cautious control (4.11) by the terms $F_\ell = .87$ in the denominator and $f_\ell = .85$ in the numerator. The denominator term in effect provides more "caution" whereas the numerator term is an additive probing effect. The term F_ℓ provides a "robustness" property in that the sensitivity of the future cost to parameter uncertainties as they appear in the controller (i.e. $\hat{b}^2(0)$) are minimized. Thus a new interpretation of the dual control is that it contains robustness and learning (via probing). These concepts are applicable to the multivariable dual controller in (3.19) - (3.21).

5. SIMULATION RESULTS

Performance was evaluated from 100 Monte Carlo runs for the following controllers where $b(0)$ was set to $\hat{b}(0)$ with covariance $P_b(0)$: 1) Cautious Controller 2) FOD 3) SOD

The above algorithms were tested for two cases:

a) Time varying case, $\hat{b}(0) = .05$, $P_b(0) = 1.0$,

   $V = .1$, $c = 1.0$, $W = .01$ and $W = .1$, $a = 0.9$

b) Constant case, with $\hat{b}(0) = .05$, $P_b(0) = 1.0$,

   $V = 0$, $c = 1.0$, $W = .01$ and $W = .1$, $a = 1.0$

Example a

Table 1 summarizes the results of the simulation runs. All three algorithms were tested on this example for two different levels of measurement noise covariance, $W = .01$ and $W = .1$. 100 Monte Carlo runs were performed each of 40 time steps. For each run, an average cost was computed over 40 time steps and then the averages over 100 runs are tabulated in Table 1 and Table 2. The tables clearly indicate that the SOD yields the least cost. The dual effect shows a larger improvement for larger measurement noise (i.e. $W = .1$). Run numbers 7 and 14 of the 100 Monte Carlo runs were selected for plotting. The cost and parameter value are plotted in Figures 1 through 4. It is evident that the second order dual improves upon the other two on the average.

Example b

In this case the true parameter was close to zero (i.e., $\hat{b}(0) = .05$) but constant. Table 2 summarizes the result. The average cost obtained by the SOD is

4

much lower than the other two. The SOD always exhibited excellent convergence whereas the other controllers performed poorly. In addition the new controller consistently avoided turn off and burst [5]. This was an important common feature in all the Monte Carlo runs. Runs 26 and 80 are plotted in Figures 5 and 6 respectively, as typical examples.

The simulation study has shown that the new dual controller improves upon the cost on the average. The magnitude of the improvement on the average appears to be relatively small for the noise levels used. However, the real advantage of the new dual controller is the improvement in those instances where the cautious controller and the FOD [1,2] yields unacceptable results. Although the FOD [1,2] shows improvement over the cautious controller, it has been found to be unacceptable at many time points.

6. CONCLUSION

A new adaptive dual control solution based upon the sensitivity functions of the expected future cost has been presented. This controller (SOD) takes into account the dual effect better by performing the second order Taylor series expansion of the expected future cost. The form of this controller is a modification of the one step cautious controller. The FOD of [1,2] did not have the denominator correction term like the present one. This adds stability to the new control design. Simulation results of a scalar model have shown the improvement obtained using the new dual algorithm.

REFERENCES

1. Bar-Shalom, Y., Mookerjee, P., and Molusis, J.A., "A Linear Feedback Dual Controller for a Class of Stochastic Systems," Proc. of National CNRS Colloquium, Developpement et Utilisation D'Outils et Modeles Mathematiques en Automatique, Analyse de Systemes et Traitment du Signal, Sept. 1982, Belle-Ile, France.
2. Mookerjee, P., "The Control of Helicopter Vibration", M.S. Thesis, Dept. of EECS, The Univ. of Connecticut, Storrs, March, 1983.
3. Feldbaum, A. A., Optimal Control Systems, New York: Academic Press, 1965.
4. Bar-Shalom, Y., "Stochastic Dynamic Programming, Caution and Probing", IEEE Trans. on Auto. Control, AC-26, 1184-1195, 1981.
5. Wittenmark, B., "An Active Suboptimal Dual Controller for Systems with Stochastic Parameters", Auto. Control Theory and Applications (Canada), Vol. 3, pp. 13-19, 1975.
6. Goodwin, G., and Payne, R., Dynamic System Identification: Experiment Design and Data Analysis, Mathematics in Science and Eng., Vol. 136, Academic Press, NY, 1977.
7. Milito, R., Padilla, C. S., Padilla, R. A. and Cadorin, D., "Dual Control Through Innovations", Proc. of the 19th IEEE Conf. on Decision and Control, Albuquerque, N. Mex., Dec. 1980.
8. Mosca, E., Rocchi, S., and Rappa, G., "A New Dual Control Algorithm", Proc. of the 1978 IEEE Conf. on Decision and Control, San Diego, Jan. 1979.
9. Sternby, J., "A Regulator for Time Varying Stochastic Systems", Proc. of the 7th International Federation of Auto. Control World Congress, Helinski, Finland, June 1978.
10. Tse, E., Bar-Shalom, Y. and Meier, L., "Wide-Sense Adaptive Dual Control for Nonlinear Stochastic Systems", IEEE Trans. Auto. Control AC-18, 98-108, April 1973.
11. Bar-Shalom, Y. and Tse, E., "Concepts and Methods in Stochastic Control", in C. T. Leondes (Ed) Control and Dynamic Systems: Advances in Theory and Applications, Vol. 12, Academic Press, NY, 1976.
12. Bellman, R., Dynamic Programming, Princeton Univ. Press, Princeton, NJ, 1957.
13. Bertsekas, D. P., Dynamic Programming and Stochastic Control, Academic Press, NY, 1976.
14. Padilla, C. S. and Cruz, J. B., "Sensitivity Adaptive Feedback with Estimation Redistribution", IEEE Trans. Automatic Control, Vol. AC-23, No. 3, June 1978.
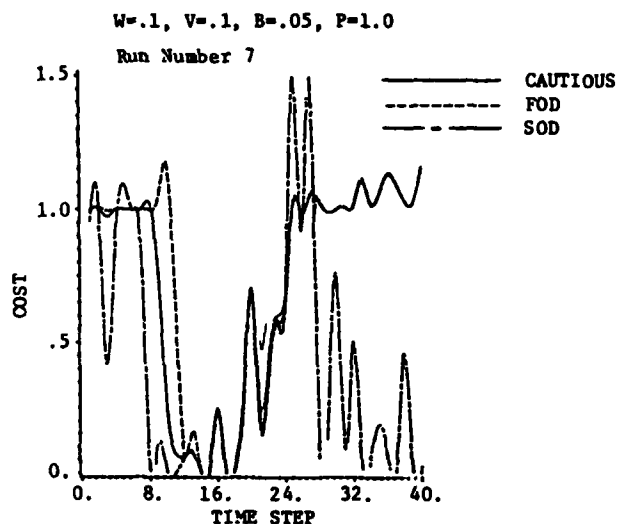
Fig. 1. Time history of cost comparing the SOD, FOD, and the cautious controller (Time varying parameter case: Run No. 7 from 100 Monte Carlo Runs).
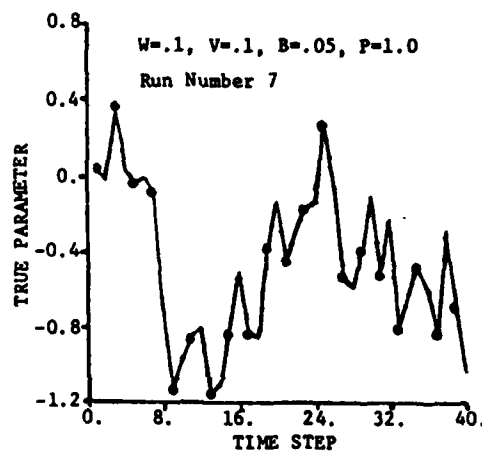

Fig. 2. Time history of parameter for Run No. 7 from the 100 Monte Carlo Runs (Time Varying Case).
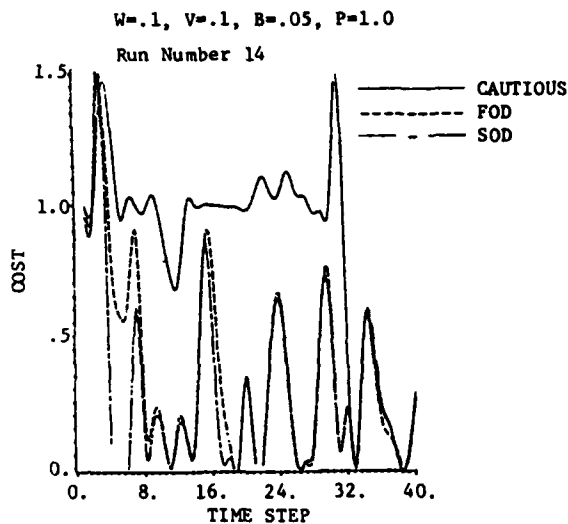
W=.1, V=.1, B=.05, P=1.0
Run Number 14

Fig. 3. Time history of cost comparing the SOD, FOD, and the cautious controller (Time varying parameter case: Run No. 14 from 100 Monte Carlo Runs)
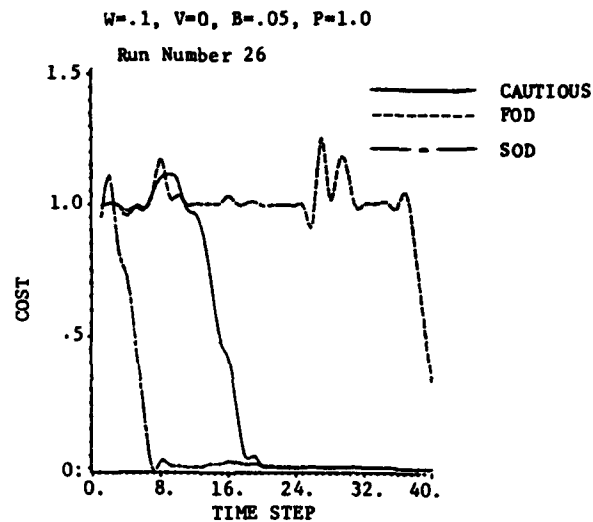
W=.1, V=0, B=.05, P=1.0
Run Number 26

Fig. 5. Time history of cost comparing the SOD, FOD, and the cautious controller (Constant parameter case: Run No. 26 from 100 Monte Carlo Runs).
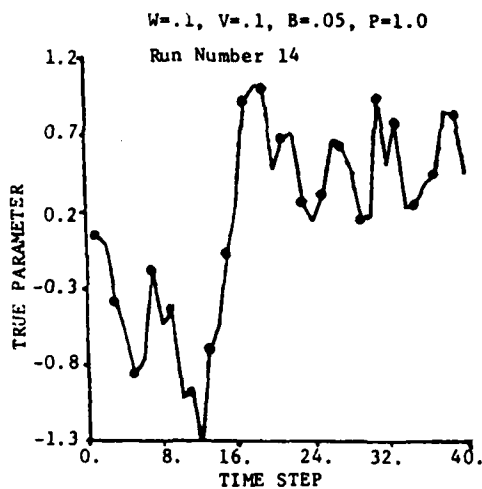
W=.1, V=.1, B=.05, P=1.0
Run Number 14

Fig. 4. Time history of parameter for Run No. 14 from 100 Monte Carlo Runs (Time Varying Case).

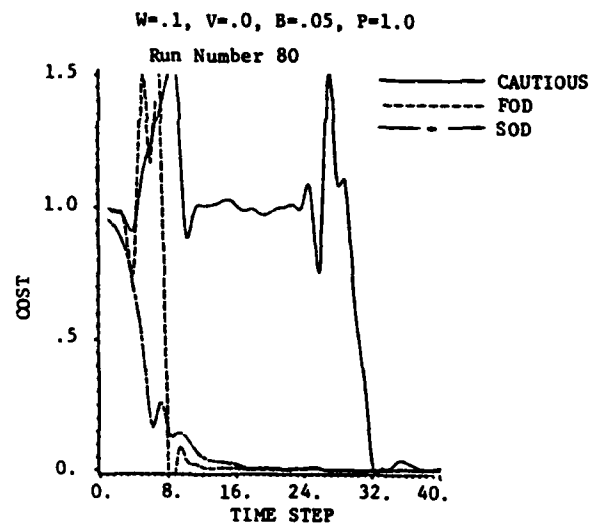W=.1, V=.0, B=.05, P=1.0
Run Number 80

Fig. 6. Time history of cost comparing the SOD, FOD, and the cautious controller (Constant parameter case: Run No. 80 from 100 Monte Carlo Runs).

Measurement Noise Covariance W	Average Cost		
	Cautious	First Order Dual	Second Order Dual
.01	.475	.469	.458
.1	.623	.608	.514

Table 1. Average Cost for the three controllers on the time varying parameter model (b(0)=.05, P_b(0)=1, V=.1, c=1)

Measurement Noise Covariance W	Average Cost		
	Cautious	First Order Dual	Second Order Dual
.01	.109	.087	.069
.1	.359	.250	.142

Table 2. Average Cost for three controllers on the Constant Parameter Model (b(0)=.05, P_b(0)=1, V=0, c=1)

END

FILMED

11-85

DTIC